

A publication of

ECOLE NATIONALE POLYTECHNIQUE (ENP)

Engineering Science

Journal

ISSN: 2716-912X

E-ISSN: 2773-4293

JUNE 2026

VOLUME 6

NUMBER 1

IN THIS NUMBER

Tinhinane Bacha, Adem Manseur, and Chems Eddine Boudjdir
Time Delay Estimation Based Integral Sliding Mode and Super-Twisting Control for Robust Robot Trajectory Tracking Pages 01 – 07

Abdelmaoula Bouaza and Rachida Touhami
Augmented Circular UWB Antenna for Dual-Band WLAN and C-Band Communication in Military Mini-Drones Pages 08 – 10

Salih Aoufi, Cherif Larbas and Aissa Chouder
A MPPT-Based ABC Algorithm Applied to PV Pumping based BLDC Motor Drive Under Partial Shading Pages 11 – 17

Fatma Zohra Hamidatou-Hamadi, Mohamed Lamine Hamidatou and Latifa Hamanni-Mitiche
FPGA Implementation for Fast and Accurate Medical Image Segmentation Pages 18 – 27

Elouahab Bouguenna, Samir Ladaci, Bilal Taghezouit, Amine Berouaken, and Walid Merrouche
SCHO-Optimized Fractional-Order Control Strategy for Enhanced Alternator Voltage Stability Pages 28 – 34

Ehsan Altayef, Sofiane Chiheb, Fateh Anayi, Michael Packianather and Eheda Hassan
Thermal Effects of Artificial Lamination Faults in a Three-Phase Transformer Core Using Infrared Thermography Pages 35 – 40

Akram Salem, Mian Hammad Nazir, Leshan Uggalla, Abdulla Rahil and Khayri Abu Sayf
Review of ISO 15118-20 with TLS1.3 for Secure and Trustworthy EV Charging Infrastructure Pages 41 – 47

Kamel Dine Remili, Adel Belouchrani, Riad Hartani and Nadjoua Boucourine
Tech Diplomacy: A Comprehensive Framework for Addressing Critical Technologies Pages 48 – 53

Noir EL Houda Benharkat, Souad Bentaalla-Kaced, and Abdelmalek Chergui
Temporal Integration Framework for Segment Level Cumulative Ergonomic Exposure Modeling Pages 54 – 64

Time Delay Estimation Based Integral Sliding Mode and Super-Twisting Control for Robust Robot Trajectory Tracking

Tinhinane Bacha, Adem Manseur, and Chems Eddine Boudjdir

Abstract—This paper improved trajectory tracking control strategy for robotic manipulators, combining Time Delay Control (TDC), Integral Sliding Mode (ISM), and the Super-Twisting Algorithm (TDC-IS-STA). The Time Delay Control framework provides a model-free aspect by estimation of unknown nonlinear dynamics and external disturbances using previously measured system data, while the integral sliding surface eliminates the reaching phase and ensures robustness from the initial instant of operation. To further mitigate the chattering phenomenon inherent in classical sliding mode control, the Super-Twisting Algorithm is incorporated to generate a continuous control signal while preserving finite-time convergence and robust performance. The proposed controller is validated through numerical simulations performed in MATLAB/Simulink on a parallel Delta robot model under multiple operating scenarios. Comparative study against control approaches demonstrate that the proposed TDC-IS-STA strategy achieves superior tracking accuracy, reduced chattering in the actuator torque signals.

Keywords—Time Delay Control, Integral Sliding Mode, Super-Twisting Algorithm, Trajectory Tracking, Delta robot.

NOMENCLATURE

PID	Proportional Integral Derivative.
TDC	Time Delay Control.
TDE	Time Delay Estimation.
SMC	Sliding Mode Control.
ISM	Integral Sliding Mode.
STA	Super-Twisting Algorithm.

I. INTRODUCTION

In modern industrial automation, achieving high precision, repeatability, and fast dynamic response has become an essential requirement for many robotic applications. Parallel robotic structures [1] have attracted considerable attention due to their superior rigidity, high positioning accuracy, and enhanced load-to-weight ratio compared to conventional serial manipulators. Owing to these advantages, parallel robots are widely used in many applications [2] such as high-speed pick-and-place operations, assembly processes, and trajectory tracking tasks that require both accuracy and rapid motion.

The Delta robot is one of the most well-known parallel robotic structures [3, 4]. Due to its lightweight moving platform and parallel kinematic architecture, the Delta robot can achieve fast,

accurate motions. Despite these advantages, controlling parallel manipulators remains challenging due to their highly nonlinear dynamics, strong coupling effects, parameter uncertainties, and sensitivity to external disturbances.

Traditional control approaches have been widely used in robotic systems because they are easy to implement and inexpensive. Among these approaches, PID have been extensively used in trajectory tracking applications [5, 6]. However, their performance tends to degrade when the robotic system operates under complex dynamics and uncertain conditions. In addition, these controllers rely on fixed tuned gains, which can become ineffective when system dynamics vary, thereby limiting tracking accuracy and robustness.

To improve trajectory tracking performance, several advanced control strategies have been proposed in the literature, including adaptive control [7], sliding mode control [8], and model-based [9]. Among these techniques, Time Delay Control (TDC) has received considerable attention due to its simple structure and reduced dependence on an accurate dynamic model. Using previously measured system data, TDC estimates unknown dynamics and disturbances, providing an efficient, model-free framework capable of achieving satisfactory tracking performance in nonlinear robotic systems.

Despite its effectiveness, the performance of TDC may deteriorate in the presence of Time Delay Estimation (TDE) errors, particularly under rapid dynamic variations and external disturbances. To overcome this limitation and enhance system robustness, TDC has been combined with various advanced control techniques [10–13], such as Sliding Mode Control (SMC). The Integral Sliding Mode (ISM) technique [14], in particular, has attracted considerable attention due to its ability to eliminate the reaching phase and ensure robustness from the initial instant of operation. By introducing an integral sliding surface, the system states are forced to evolve on the sliding manifold throughout

Manuscript received June 6, 2026; accepted June 28, 2026.

T. Bacha and A. Manseur are affiliated with the Department of Automatic, University of Science and Technology Houari Boumediene, Algiers, Algeria (bachatinhinane06@gmail.com; ademmanseur@gmail.com)

C. Boudjdir is affiliated with LRPE laboratory, University of science and technology Houari Boumediene, Algiers, Algeria (chemseddine.boudjdir@usthb.edu.dz)

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.381

the motion, resulting in improved disturbance rejection and enhanced tracking accuracy.

However, classical SMC still suffers from the chattering phenomenon caused by discontinuous switching control actions [15, 16]. This phenomenon produces high-frequency oscillations in the control signal that may lead to mechanical wear, actuator degradation, unwanted vibrations, and excitation of unmodeled dynamics.

Several techniques have been proposed in the literature to reduce chattering, including boundary layer methods and saturation functions [16, 17]. While these methods can reduce high-frequency oscillations, they often decrease robustness and tracking performance. To overcome this limitation, the Super-Twisting Algorithm (STA), which belongs to the class of higher-order sliding mode techniques has emerged as an effective solution for chattering mitigation [18, 19]. Unlike classical SMC, the STA generates a continuous control signal while maintaining the robustness of sliding mode control. Furthermore, STA guarantees finite-time convergence of the sliding variable and its derivative, providing effective chattering mitigation. Consequently, combining integral sliding surfaces with the STA provides enhanced disturbance rejection, reduced chattering, and higher trajectory tracking accuracy.

Motivated by these advantages, this work proposes a robust model-free control strategy combining Time-Delay Control, Integral Sliding Mode, and the Super-Twisting Algorithm (TDC-IS-STA) for robotic manipulator trajectory tracking. The proposed approach does not require an accurate dynamic model and can be applied to robotic manipulators with arbitrary degrees of freedom (n-DOF). The integral sliding surface is introduced to improve disturbance rejection from the initial instant and enhance tracking precision, while the STA component is employed to suppress chattering and maintain robust performance in the presence of uncertainties and TDE errors. The proposed controller aims to achieve accurate and smooth trajectory tracking while ensuring robustness and stability under different operating conditions. Furthermore, the asymptotic stability of the closed-loop system is rigorously established using Lyapunov stability theory. In addition, a comparative study using a similar control approach is conducted on a parallel Delta robot to further demonstrate the effectiveness and applicability of the proposed method.

The remainder of this paper is organized as follows. Section II presents the fundamental concepts of TDC and ISM. Section III describes the proposed TDC-IS-STA control strategy and discusses its stability and convergence analysis. Section IV presents the simulation results and performance evaluation. Finally, Section V concludes the paper.

II. BRIEF REVIEW OF TIME-DELAY CONTROL BASED SUPER-TWISTING SLIDING MODE CONTROLLER

A. Time-Delay Control

Time Delay Control is a robust nonlinear control technique used to compensate unknown dynamics and external disturbances in uncertain systems by using information from a previous sampling instant. The main principle of TDC is that the nonlinear dynamics of the system do not change significantly during a

very short time interval. Therefore, the unknown dynamics at the current time can be approximated by their delayed values. For an n-degree-of-freedom (n-DOF) robot manipulator, the dynamic model is expressed as:

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + d(t) = \tau \quad (1)$$

Where $q \in \mathbb{R}^n$ denotes the vector of joint positions, while $M(q)$ corresponds to the inertia matrix of the manipulator. The term $C(q, \dot{q})\dot{q}$ describes the Coriolis and centrifugal forces, and $G(q)$ represents the gravitational effects acting on the robot. In addition, $d(t)$ includes external disturbances and model uncertainties, and τ defines the vector of control torques applied to the joints.

The dynamic behavior of the n -DOF robotic manipulator can be reformulated by introducing a nominal constant inertia matrix $\bar{M}$. By separating the nominal and uncertain parts of the inertia matrix, the system dynamics become

$$\bar{M}\ddot{q} + (M(q) - \bar{M})\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + d(t) = \tau \quad (2)$$

in which $\bar{M}$ is selected as a constant diagonal positive-definite matrix. Based on this decomposition, the manipulator dynamics can be rewritten in a compact representation as

$$\ddot{q}_t = \bar{M}^{-1}\tau_t + N_t \quad (3)$$

with the nonlinear uncertain term N_t defined by

$$N_t = -\bar{M}^{-1} [(M(q_t) - \bar{M})\ddot{q}_t + C(q_t, \dot{q}_t)\dot{q}_t + G(q_t) + d(t)]$$

The index t refers to the current sampling instant. Since the exact nonlinear dynamics are generally difficult to obtain, the unknown term N_t is approximated using the (TDE) principle. Assuming that the system dynamics vary slowly over a sufficiently small delay interval L , the estimated nonlinear term can be expressed as

$$\hat{N}_t \approx N_{t-L} \quad (4)$$

which yields

$$\hat{N}_t = \ddot{q}_{t-L} - \bar{M}^{-1}\tau_{t-L} \quad (5)$$

The tracking objective is achieved by forcing the manipulator joints to follow a desired reference trajectory q_d . Accordingly, the position and velocity tracking errors are introduced as

$$\tilde{q}_t = q_{d,t} - q_t \quad (6)$$

and

$$\dot{\tilde{q}}_t = \dot{q}_{d,t} - \dot{q}_t \quad (7)$$

respectively, the estimation accuracy of the TDE approach is characterized through the time-delay estimation error defined by

$$\tilde{N} = \hat{N}_t - N_t \quad (8)$$

B. Integral sliding surface

The integral sliding surface is a sliding manifold that incorporates both the tracking error and its integral, ensuring improved robustness against system uncertainties and disturbances while eliminating the reaching phase. The integral sliding surface is defined as

$$s_t = \dot{\tilde{q}}_t + K_d\tilde{q}_t + K_p \int_0^t \tilde{q}_\tau d\tau \quad (9)$$

in which K_p and K_d denote positive definite gain matrices associated with the position and velocity tracking errors, respectively. Since the manipulator is initialized at the desired position, i.e.,

$$q(0) = q_d(0) \quad (10)$$

the initial tracking error is zero, $e(0) = 0$. Consequently, the sliding variable satisfies

$$s(0) = 0 \quad (11)$$

which ensures that the reaching phase will be eliminated.

III. PROPOSED CONTROL LAW AND STABILITY ANALYSIS

A. Control Design

The proposed TDC-IS-STA is given by :

$$\tau_t = \tau_{t-L} - \bar{M}\ddot{q}_{t-L} + \bar{M}(\ddot{q}_{d,t} + K_d\dot{q}_t + K_p\tilde{q}_t + u_{st}) \quad (12)$$

The terms τ_{t-L} and $\ddot{q}_{t-L}$ correspond to the delayed control input and delayed joint acceleration used in the Time Delay Control framework.

To enhance the robustness of the closed-loop system and attenuate the chattering effect, a Super-Twisting term u_{st} is incorporated into the control law.

The super-twisting algorithm is expressed as

$$u_{st} = k_1|s_t|^{1/2} \text{sgn}(s_t) + v \quad (13)$$

with the auxiliary variable v governed by

$$\dot{v} = k_2 \text{sgn}(s_t) \quad (14)$$

B. proof of convergence

1— Proposition: For the n-DOF robotic manipulator described by (1), the tracking error converges asymptotically to zero under the proposed TDC-IS-STA control law defined in (12), thereby ensuring asymptotic tracking stability of the closed-loop system.

2— Proof: Differentiating (9) with respect to time yields

$$\dot{s}_t = \ddot{q}_t + K_d\dot{q}_t + K_p\tilde{q}_t \quad (15)$$

Since

$$\ddot{q}_t = \ddot{q}_{d,t} - \ddot{q}_t \quad (16)$$

then

$$\dot{s}_t = \ddot{q}_{d,t} - \ddot{q}_t + K_d\dot{q}_t + K_p\tilde{q}_t \quad (17)$$

Substituting (3) gives

$$\dot{s}_t = \ddot{q}_{d,t} + K_d\dot{q}_t + K_p\tilde{q}_t - N_t - \bar{M}^{-1}\tau_t \quad (18)$$

Multiplying the control law (12) by $\bar{M}^{-1}$ yields

$$\bar{M}^{-1}\tau_t = \bar{M}^{-1}\tau_{t-L} - \ddot{q}_{t-L} + \ddot{q}_{d,t} + K_d\dot{q}_t + K_p\tilde{q}_t + u_{st} \quad (19)$$

Using (5), (19) becomes

$$\bar{M}^{-1}\tau_t = -\hat{N}_t + \ddot{q}_{d,t} + K_d\dot{q}_t + K_p\tilde{q}_t + u_{st} \quad (20)$$

Substituting (20) into (18) leads to

$$\dot{s}_t = \hat{N}_t - N_t - u_{st} \quad (21)$$

using the TDE estimation error given by (8), one obtains

$$\dot{s}_t = \tilde{N} - u_{st} \quad (22)$$

Substituting (13) and (14) into (22) gives

$$\dot{s}_t = -k_1|s_t|^{1/2} \text{sgn}(s_t) - v - \int_0^t k_2 \text{sgn}(s_t) + \tilde{N} \quad (23)$$

By introducing the following state variables,

$$z_1 = s_t, \quad (24)$$

$$z_2 = -K_2 \int_0^t \text{sign}(s_t) d\tau + \tilde{N} \quad (25)$$

the system dynamics can be expressed as follows

$$\dot{z}_1 = -k_1|z_1|^{1/2} \text{sign}(z_1) + z_2, \quad (26)$$

$$\dot{z}_2 = -k_2 \text{sign}(z_1) + \dot{\tilde{N}} \quad (27)$$

Defining

$$\zeta = \begin{bmatrix} |z_1|^{1/2} \text{sign}(z_1) \\ z_2 \end{bmatrix} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \quad (28)$$

the subsequent expressions are derived as

$$\dot{\zeta}_1 = |z_1|^{-1/2} \left[-\frac{1}{2}k_1|z_1|^{1/2} \text{sign}(z_1) + \frac{1}{2}z_2 \right] \quad (29)$$

$$\dot{\zeta}_2 = |z_1|^{-1/2} \left[-k_2|z_1|^{1/2} \text{sign}(z_1) \right] + \dot{\tilde{N}} \quad (30)$$

The aggregated uncertainty term $\tilde{N}$ is assumed to be continuously differentiable, and there exists a positive scalar δ_N satisfying

$$|\dot{\tilde{N}}| \leq \delta_N, \quad (31)$$

which is globally bounded.

$$\begin{aligned} \dot{\zeta} &= \begin{bmatrix} \dot{\zeta}_1 \\ \dot{\zeta}_2 \end{bmatrix} \\ &= -|z_1|^{-1/2} \begin{bmatrix} \frac{1}{2}k_1 & -\frac{1}{2} \\ k_2 & 0 \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \dot{\tilde{N}} \end{bmatrix} \end{aligned} \quad (32)$$

$$\dot{\zeta} = -|z_1|^{-1/2} A \zeta - \begin{bmatrix} 0 \\ \dot{\tilde{N}} \end{bmatrix} \quad (33)$$

where

$$A = \begin{bmatrix} \frac{1}{2}k_1 & -\frac{1}{2} \\ k_2 & 0 \end{bmatrix} \quad (34)$$

Consider the following Lyapunov function candidate

$$V_\zeta = \frac{1}{2} \zeta^T P \zeta, \quad (35)$$

where P denotes a positive definite matrix given by

$$P = \frac{1}{2} \begin{bmatrix} 4k_2 + k_1^2 & -k_1 \\ -k_1 & 2 \end{bmatrix} \quad (36)$$

It is positive definite if $k_2 > 0$ Taking the time derivative of V_ζ yields

$$\dot{V}_\zeta = \frac{1}{2} \dot{\zeta}^T P \zeta + \frac{1}{2} \zeta^T P \dot{\zeta} \quad (37)$$

Substituting $\dot{\zeta}$ into the previous equation gives

$$\dot{V}_\zeta = -|z_1|^{-1/2} \zeta^T P A \zeta + \zeta^T P \begin{bmatrix} 0 \\ \dot{N} \end{bmatrix} \quad (38)$$

Substituting P

$$\dot{V}_{\zeta 1} = -|z_1|^{-1/2} \zeta^T Q \zeta + \dot{N} \left[-\frac{k_1}{2} \ 1 \right] \zeta \quad (39)$$

$$\dot{V}_\zeta \leq -|z_1|^{-1/2} \zeta^T Q_\delta \zeta \quad (40)$$

where

$$Q = A^T P + P A = \frac{k_1}{2} \begin{bmatrix} 2k_2 + k_1^2 & -k_1 \\ -k_1 & 1 \end{bmatrix} \quad (41)$$

and

$$Q_\delta = \frac{k_1}{2} \begin{bmatrix} 2k_2 + k_1^2 - \delta_N & -k_1 - \frac{\delta_N}{k_1} \\ -k_1 - \frac{\delta_N}{k_1} & 1 \end{bmatrix} \quad (42)$$

if the gains satisfy

$$k_1 > 0, \quad k_2 > \frac{\delta_N(\delta_N + 3k_1^2)}{2k_1^2}, \quad (43)$$

Then, it follows that

$$z_1 \rightarrow 0, \quad z_2 \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty,$$

which implies that

$$s_t \rightarrow 0, \quad \tilde{q}_t \rightarrow 0 \quad \text{asymptotically as} \quad t \rightarrow \infty.$$

IV. SIMULATION RESULTS AND DISSCUSION

In this section, the effectiveness of the proposed control strategy is investigated through numerical simulations carried out in MATLAB/Simulink on a parallel Delta robot model shown in Fig. 1. The Delta robot represents a suitable platform for validating advanced trajectory tracking controllers due to its highly nonlinear coupled dynamics and its capability to operate at high speeds [3]. Detailed formulations of the inertia matrix M , Coriolis and centrifugal matrix C , and gravity vector G are available in [20].

A. Simulation Setup

The reference trajectory adopted for the simulations follows a semi-elliptical path. The motion starts from the point $(0.25, 0, -0.43)$ and ends at $(-0.25, 0, -0.43)$. To demonstrate the performance of the proposed controller, referred to as TDC+IS+STA, comparative analyses are performed with two additional control approaches: the controller introduced in (3) and the Time Delay Estimation based Integral Sliding controller (TDE+IS) described by Equation (44). The comparison focuses on the tracking capability and robustness of each control strategy under identical operating conditions.

The control input associated with the TDE+IS approach can be written as

$$\tau_t = -\bar{M} \ddot{q}_{t-L} + \tau_{t-L} + \bar{M} (\ddot{q}_d + K_d \dot{\tilde{q}}_t + K_p \tilde{q}_t + K_s \text{sgn}(s_t)) \quad (44)$$

while K_p , K_d , and K_s represent positive control gains related to the proportional, derivative, and switching actions, respectively.

The simulations are presented under three distinct scenarios.

B. Scenario 1: Nominal Trajectory Tracking

The first simulation case focuses on nominal operating conditions in order to analyze the tracking capability of the different controllers under moderate motion requirements. In this scenario, the desired joint trajectory is executed within $t = 0.8$ s. In the TDE+IS scheme, the saturation function is employed in place of the signum function within the integral sliding surface in order to smooth the control input and avoid the chattering phenomenon. Several simulation tests were performed in order to tune the controller parameters and achieve satisfactory tracking performance. Based on these numerical experiments, The optimal set of control gains is shown in Table 1.

Table. I
CONTROLLER PARAMETERS

Controller	Parameters	Values
TDC+IS	K_p, K_d, K_s M	1800, 1800, 1460 0.0003
TDC+IS+STA	K_p, K_d, k_1, k_2 M	1800, 1800, 700, 100 0.0003

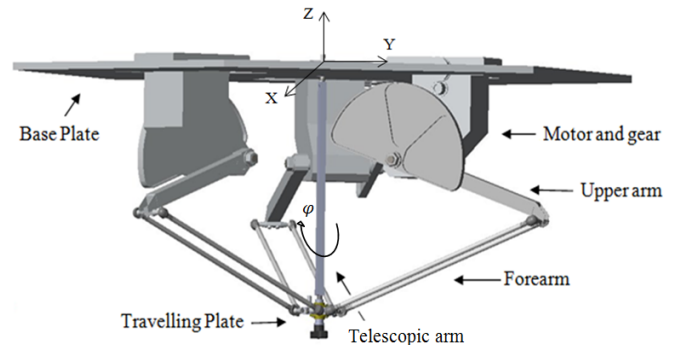
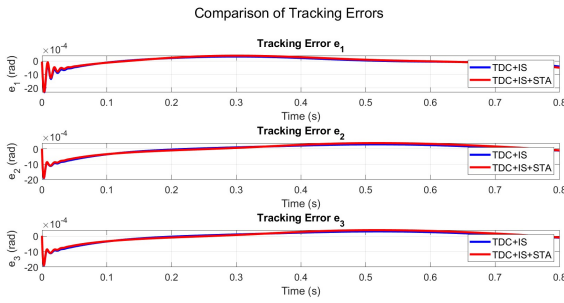


Fig. 1: Delta robot structure

Fig. 2: Trajectory tracking in the task space**Fig. 3:** Tracking errors in scenario 1 for each joint using TDC+IS and TDC+IS+STA.

The simulation results presented in Fig. 2, Fig. 3 and Fig. 4 indicate that all considered controllers are capable of achieving satisfactory trajectory tracking while maintaining relatively smooth control torques. The relatively small performance improvement observed in Scenario 1 is expected because it involves a low-cadence, slow trajectory, which imposes relatively mild dynamic demands on the controllers. However, the proposed control strategy provides superior tracking precision compared with the other approaches. This improvement is confirmed by the lower Root Mean Square (RMS) tracking errors reported in Table II.

C. Scenario 2: High-Speed Motion Condition

The second simulation scenario investigates a more aggressive operating condition intended to assess the robustness and dynamic response of the compared control methods under fast trajectory execution. For this case, the desired joint motion is completed within a shorter duration of $t = 0.2$ s. Such rapid movements generate stronger transient effects and increase the difficulty of the tracking task due to the highly nonlinear dynamics of the robotic system. As illustrated in Fig. 5, the tracking errors of the TDC+IS+STA controller remain bounded and converge faster compared to the standard TDC+IS approach, demonstrating improved robustness under high-cadence operating conditions. Furthermore, the control torque profiles shown in Fig. 6 confirm that the TDC+IS+STA controller produces smoother torque signals with reduced chattering, which is particularly advantageous for practical implementation on real robotic systems.

D. Scenario 3: Composite Semi-Elliptical Three-Dimensional Trajectory

In the third simulation study, the controllers are evaluated using a more complex three-dimensional composite semi-elliptical trajectory, as illustrated in Fig. 7. This scenario is designed to examine the capability of the proposed approach to accurately track spatial motions involving simultaneous variations along multiple axes. The objective of this test is to validate the effectiveness of the control strategy for realistic robotic operations requiring smooth and precise 3D trajectory tracking under nonlinear coupled dynamics. As illustrated in Fig. 8, the tracking errors of the TDC+IS+STA controller remain bounded and exhibit faster convergence compared to the standard TDC+IS approach, particularly at the trajectory transition point, demon-

Fig. 4: Control torque in scenario 1 for each joint using TDC+IS and TDC+IS+STA.**Fig. 5:** Tracking errors in scenario 2 for each joint using TDC+IS and TDC+IS+STA.

strating improved robustness under complex spatial motion conditions. Furthermore, the control torque profiles shown in Fig. 9 confirm that the TDC+IS+STA controller produces smoother torque signals with significantly reduced chattering across all three joints, which is particularly advantageous for practical implementation on real robotic systems.

RMS TRACKING ERRORS FOR THE COMPARED CONTROLLERS

Table. II

RMS TRACKING ERRORS FOR SCENARIO 1 (rad)

Controller	Joint 1	Joint 2	Joint 3
TDC+IS	3.463×10^{-4}	3.587×10^{-4}	3.587×10^{-4}
TDC+IS+STA	3.423×10^{-4}	3.583×10^{-4}	3.583×10^{-4}

Table. III

RMS TRACKING ERRORS FOR SCENARIO 2 (rad)

Controller	Joint 1	Joint 2	Joint 3
TDC+IS	0.0056	0.0069	0.0069
TDC+IS+STA	0.0044	0.0047	0.0047

Table. IV

RMS TRACKING ERRORS FOR SCENARIO 3 (rad)

Controller	Joint 1	Joint 2	Joint 3
TDC+IS	7.265×10^{-4}	1.622×10^{-4}	1.608×10^{-4}
TDC+IS+STA	5.592×10^{-4}	1.776×10^{-4}	1.752×10^{-4}

V. CONCLUSION

This paper presented a robust model-free trajectory tracking control strategy based on the combination of Time Delay Control, Integral Sliding Mode, and the Super-Twisting Algorithm for robotic manipulators. The proposed TDC-IS-STA approach enhances tracking accuracy while reducing the chattering phenomenon commonly associated with conventional sliding mode methods. By integrating the integral sliding surface and the super-twisting algorithm within the TDC framework, the controller ensures improved robustness against parameter uncertainties, external disturbances, and TDE errors. In addition, a Lyapunov stability analysis was conducted to guarantee the stability and convergence of the closed-loop system. The effectiveness of the proposed method was investigated on a parallel Delta robot under different operating conditions and complex three-dimensional trajectories. Comparative results demonstrated superior tracking performance, smoother control signals, and enhanced robustness compared to existing approaches. These results confirm the suitability of the proposed strategy for high-performance robotic applications involving nonlinear and uncertain dynamics.

Fig. 6: Control torque in scenario 2 for each joint using TDC+IS and TDC+IS+STA.

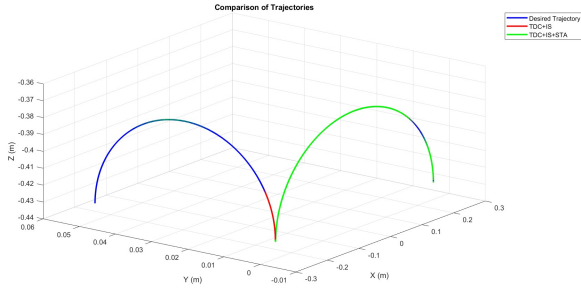


Fig. 7: trajectory tracking in the task space for scenario 3

REFERENCES

- [1] H. D. Taghirad, *Parallel Robots: Mechanics and Control*, 2nd ed. Boca Raton, FL: CRC Press, 2026.
- [2] Z. Zhang, Z. Shao, Z. You, X. Tang, B. Zi, G. Yang, C. Gosselin, and S. Caro, "State-of-the-art on theories and applications of cable-driven parallel robots," *Frontiers of Mechanical Engineering*, vol. 17, no. 3, p. 37, 2022.
- [3] N. Cretescu, M. Neagoe, and R. Saulescu, "Dynamic analysis of a delta parallel robot with flexible links and joint clearances," *Applied Sciences*, vol. 13, no. 11, p. 6693, 2023.
- [4] I. Bonev, "Delta parallel robot—the story of success," *ParallelMIC*, May 2001.
- [5] T. Häggglund and J. L. Guzmán, "Give us PID controllers and we can control the world," *IFAC PapersOnLine*, vol. 58, no. 7, pp. 103–108, 2024.
- [6] S. Amrane, C. E. Boudjedir, and D. Boukhetala, "Fractional-order pd control of a parallel delta robot: Experimental results," *ENP Engineering Science Journal*, vol. 5, no. 2, pp. 7–11, 2025.
- [7] L. Wang, Z. Lu, X. Liu, K. Liu, and D. Zhang, "Adaptive control of a parallel robot via backstepping technique," *International Journal of Systems, Control and Communications*, vol. 1, no. 3, pp. 312–324, 2009.
- [8] A. Kihal, C. E. Boudjedir, B. Talbi, A. Sahli, and A. Krama, "Adaptive model-free sliding mode control for boost converter with flexibility to varying operating conditions," *Journal of Electrical Engineering*, vol. 77, no. 1, pp. 77–88, 2026.
- [9] J. A. Rossiter, *Model-Based Predictive Control: A Practical Approach*, ser. CRC Press Control Series. Boca Raton, FL: CRC Press, 2003.
- [10] Y. Wang, J. Chen, K. Zhu, B. Chen, and H. Wu, "Time-delay control of cable-driven robots with adaptive fractional-order nonsingular terminal sliding mode," *IEEE Access*, vol. 7, pp. 157 716–157 726, 2019.
- [11] S. Han, H. Wang, Y. Tian, and N. Christov, "Time-delay estimation based computed torque control with robust adaptive RBF neural network compensator for a rehabilitation exoskeleton," *ISA Transactions*, vol. 97, pp. 171–181, 2020.
- [12] C. E. Boudjedir, "A widely-practical model-free prescribed time control for trajectory tracking of n-dof robot manipulators," *European Journal of Control*, p. 101396, 2025.
- [13] S. Kim and J.-H. Suh, "A study on robust control scheme using prescribed performance-based time delay control and rbf neural network," *IEEE Access*, vol. 13, pp. 180 513–180 522, 2025.
- [14] C. E. Boudjedir, M. Bouri, and D. Boukhetala, "An enhanced adaptive time delay control-based integral sliding mode for trajectory tracking of robot manipulators," *IEEE Transactions on Control Systems Technology*, vol. 31, no. 3, pp. 1042–1050, 2023.
- [15] A. Rehan, I. Boiko, and Y. Zweiri, "Chaotic chattering in sliding mode control systems," *IEEE Transactions on Automatic Control*, vol. 69, no. 11, pp. 7925–7931, 2024.
- [16] J. Guldner and V. I. Utkin, "The chattering problem in sliding mode systems," in *Proceedings of the 14th International Symposium of Mathematical Theory of Networks and Systems (MTNS)*, Perpignan, France, June 2000.
- [17] A. Sahamijoo, F. Piltan, M. H. Mazloom, M. R. Avazpour, H. Ghiasi, and N. B. Sulaiman, "Methodologies of chattering attenuation in sliding mode controller," *International Journal of Hybrid Information Technology*, vol. 9, no. 2, pp. 11–36, 2016.
- [18] H. Saied, A. Chemori, M. Bouri, M. E. Rafei, and C. Francis, "Feedforward super-twisting sliding mode control for robotic manipulators: Application to pkms," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 3167–3184, 2023.
- [19] I.-L. G. Borlaug, K. Y. Pettersen, and J. T. Gravdahl, "The generalized super-twisting algorithm with adaptive gains," *International Journal of Robust and Nonlinear Control*, vol. 32, no. 13, pp. 7240–7270, 2022.
- [20] C. E. Boudjedir, M. Bouri, and D. Boukhetala, "Iterative learning control for trajectory tracking of a parallel delta robot," *at-Automatisierungstechnik*, vol. 67, no. 2, pp. 145–156, 2019.

VI. BIOGRAPHY

AUTHORS' BIOGRAPHIES

Bacha Tinhinane received the B.Sc. degree in Automation Engineering from the University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria, in 2024. She

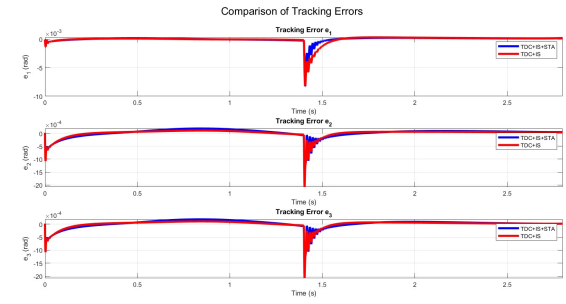


Fig. 8: Tracking errors in scenario 3 for each joint using TDC+IS and TDC+IS+STA.

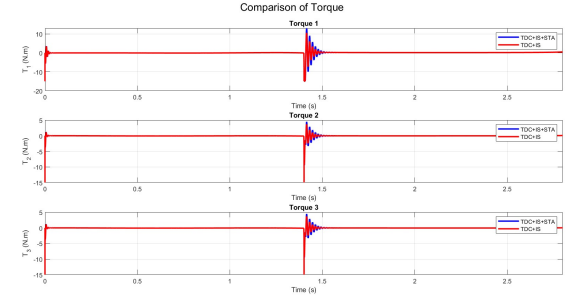


Fig. 9: Control torque in scenario 3 for each joint using TDC+IS and TDC+IS+STA.

is currently pursuing the M.Sc. degree in Control Engineering at USTHB. her research interests include robotic systems, non-linear control, sliding mode control, time-delay control, neural networks.

Adem Manseur received the B.Sc. degree in Automation Engineering from the University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria, in 2024. He is currently pursuing the M.Sc. degree in Control Engineering at USTHB. His research interests include robotic systems, non-linear control, sliding mode control, time-delay control, neural networks.

Chems Eddine Boudjedir received the M.Sc. and Ph.D. degrees in control engineering from the ole Nationale Polytechnique, Algiers, Algeria, in 2015 and 2019, respectively. He is currently an Associate Professor with the University of Science and Technology Houari Boumediene, Algiers. Hisresearch interests include iterative learning control, sliding mode control, time delay control, disturbance observer, and robotics.

Augmented Circular UWB Antenna for Dual-Band WLAN and C-Band Communication in Military Mini-Drones

Abdelmaoula Bouaza and Rachida Touhami

Abstract—This paper presents a novel UWB circular microstrip antenna (UWB-CMSA) designed for handheld devices with a partial ground plane, achieving resonance from 1.5 GHz to 11.2 GHz. The antenna effectively covers various bands including GPS (1.575 GHz), DCS (1.71 - 1.88 GHz), PCS (1.85–1.99 GHz), UMTS (1.925 - 2.175 GHz), Bluetooth (2.4 GHz), LTE (2.5/2.69 GHz), WLAN (2.4/3.5/5 GHz), WiMAX (2.5/3.5/5.5 GHz), and X-band applications (8–12 GHz) encompassing satellite communication and military radar. The antenna exhibits good gain values (ranging from 2.45 dB to 5.5 dB) across these desired frequency bands and maintains high efficiency. The design was validated through fabrication and Vector Network Analyzer (VNA) measurements, demonstrating excellent agreement with simulated results. Uniquely, the antenna is shaped like the emblem of the Algerian Civil Protection Agency, allowing for seamless integration into their equipment (drones, communication devices, fire trucks, etc.). This integration enables diverse applications in the 1.5 - 11.2 GHz range, including identification, localization, and environmental monitoring in critical situations. Furthermore, the proposed design incorporates the Civil Protection emblem for potential performance enhancement when mounted on mini drones. CST software was employed for design and simulation purposes.

Keywords— UWB, GPS, WiMAX, X-band, WLAN, VNA.

I. INTRODUCTION

Modern wireless communication systems utilize multiple frequency bands to achieve high data rates. This necessitates antennas with wide operating bandwidths. Multiband, wideband, and ultra-wideband (UWB) antennas accommodate diverse applications within a single antenna system. Cost-effectiveness is crucial; ideally, the chosen antenna (multiband or UWB) can be seamlessly integrated with the communication system. [1]. The emergence of UWB technology has captivated academia and industry since the Federal Communications Commission (FCC) allocated the 3.1–10.6 GHz spectrum for UWB communication applications in 2002 [2]. This extensive spectrum allocation has spurred significant interest in developing UWB technology for short-range wireless communication, imaging radar, remote sensing, and location confirmation applications.

UWB antenna design prioritizes three key objectives: high gain, wide operating bandwidth, and good radiation efficiency, all while maintaining efficient manufacturing processes. UWB antennas offer a distinct advantage over traditional narrowband systems—compact size and low profile [3]. Additionally, planar UWB antennas are favored for their thin profile, lightweight construction, and ease of manufacturing, making them well-suited for multimode communication systems [4].

Research has explored various antenna shapes for achieving UWB characteristics, including rectangular, circular, elliptical, and curved monopoles, as reported in [5–8]. Microstrip patch

antennas offer design flexibility and can be realized in various shapes such as dipoles, squares, rectangles, triangles, circles, circular rings, ring sectors, and disk sectors. Circular patches are particularly advantageous due to their design flexibility, offering the highest bandwidth in GHz, acceptable loss characteristics, enhanced gain, and desirable electric and magnetic field patterns [9].

The integration of well-known logos into the design of patch antennas has recently attracted significant attention from researchers. This approach enables the antenna to serve a dual function: acting as a radiating element for communication over multiple frequency bands, while simultaneously maintaining the appearance of a conventional logo. Such logo-based antennas are particularly advantageous in military applications, where the logo is often already present on uniforms or equipment, ensuring visual discretion and reducing the likelihood of the antenna being identified, thereby enhancing camouflage capabilities [10–12].

Augmented antennas, or multifunctional antennas, are antennas that, in addition to their traditional functions of transmitting and receiving electromagnetic waves, integrate one or more additional features designed to enhance their performance and meet specific requirements [10]. These antennas may be tailored to function as Rectennas [13], sensor antennas [14], or aesthetically integrated graphic antennas [10–12].

This work presents the design, fabrication, and characterization of a novel circular Ultra-Wideband (UWB) patch antenna operating from 1.5 GHz to 11.2 GHz. This bandwidth encompasses both Wireless Local Area Networks (WLAN) and C-band applications. The antenna's design is inspired by the Algerian Civil Protection Agency emblem, facilitating seamless integration with their equipment, such as drones, communication devices, and fire trucks. This integration enables a variety of functionalities within the 1.5 – 11.2 GHz range, including identification, localization, and environmental monitoring during critical situations. The proposed antenna demonstrates satisfactory return loss (S_{11}), gain, and radiation

Manuscript received December 3, 2025; Accepted June 28, 2026.

Abdelmaoula Bouaza and Rachida Touhami are with the Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: abdelmaoula.bouaza@g.enp.edu.dz; rachida.touhami@g.enp.edu.dz).

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.356

efficiency. For potential use in military applications, the antenna can be mounted on a mini drone.

II. ANTENNA DESIGN

A. EVOLUTION OF THE SINGLE ELEMENT DESIGN

The fundamental antenna structure, designed using equations from references [1-5], is illustrated in Figure 1. The antenna is fabricated on FR-4 substrate material with a dielectric constant of 4.3, a loss tangent of 0.025, and a thickness of 1.6 mm, as shown in Figure 2. This configuration, known as a Circular Monopole Slot Antenna (CMSA), exhibits UWB characteristics from 2.4 GHz to 12 GHz, encompassing the essential frequency bands for various handheld device applications. Notably, the CMSA geometry serves as a foundational structure for further design enhancements.

$$F_{mn} = \left(\frac{cx_{mn}}{2\pi\sqrt{\epsilon_{eff}}} \right) \quad (1)$$

where F_{mn} is the resonant frequency, c is the speed of light in free space, x_{mn} is the first-order Bessel function root, r is the radius of the patch, and ϵ_{eff} is the effective dielectric constant.

In a circular antenna, the TM₁₁ mode is dominant and is given refer to (4), for ϵ_{eff} refers to (2). Equation (3) gives the effective radius (r_e) considered due to fringing fields [15-16]. Equation (5) gives the width of the feed.

$$\epsilon_{eff} = \left(\frac{\epsilon_r + 1}{2} \right) + \left(\frac{\epsilon_r - 1}{2} \right) \left(1 + \frac{12h}{w} \right)^{-\frac{1}{2}} \quad (2)$$

$$r_e = r \left(1 + \frac{2h}{\pi r \epsilon_r} \left(\ln \frac{\pi r}{2h} \right) + 1.7726 \right) \quad (3)$$

$$F_{mn} = \left(\frac{1.8412c}{2\pi r_e \sqrt{\epsilon_{eff}}} \right) \quad (4)$$

$$w_f = \frac{7.48h}{e^{\left(\frac{Z_0 \sqrt{\epsilon_r + 1.41}}{87} \right)}} - 1.25 \text{ T} \quad (5)$$

where ϵ_r is relative permittivity, h is substrate height, w is the width, Z_0 is the single-ended impedance, and T is the trace thickness.

The evolution steps of the creation of the antenna are shown in Figure 1.

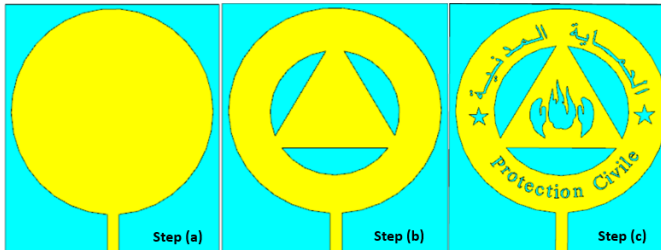


Fig. 1: Evolution steps of the proposed antenna

The structure of the proposed antenna is depicted in Figure 2, and all parameters of the proposed antenna are: $L_s = 65$ mm, $W_s = 56$ mm, $L_g = 11$ mm, $L_f = 11$ mm, $W_f = 3$ mm, $a = 28$ mm, $b = 15$ mm, and $c = 27.5$ mm.

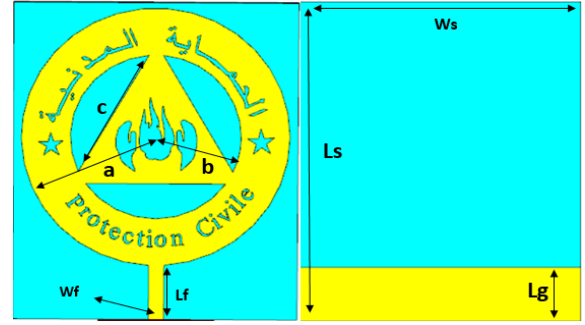


Fig. 2: Structure of proposed antenna: (a) Front view (b) Back view.

The initial design (Step 1) consisted of a simple circular monopole with a partial ground plane. This configuration produced a single two-wide band ranging from 1.7 GHz to 4 GHz. To achieve the UWB, a triple semi-circular slot was incorporated on the patch in Step 2. This element increased the current path length, resulting in UWB from 1.7 GHz to 11.2 GHz. Step 3 involved adding the shape of the Algerian Civil Protection Agency. This modification maintains the same frequency bandwidth.

B. SIMULATION RESULTS

The evolutionary stages of the proposed antenna design are presented in Figure 1. The simulated S_{11} results, obtained throughout the design process, are displayed in Figure 3. The simulated gain and radiation efficiency are shown in Figure 4.

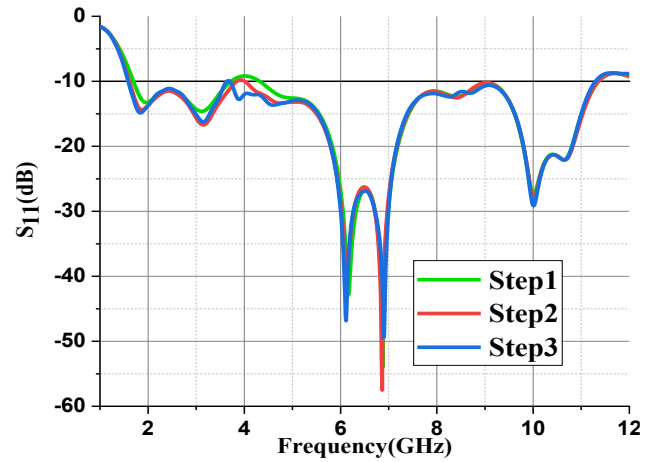


Fig. 3: Simulated S_{11} of the evolution steps of the proposed antenna

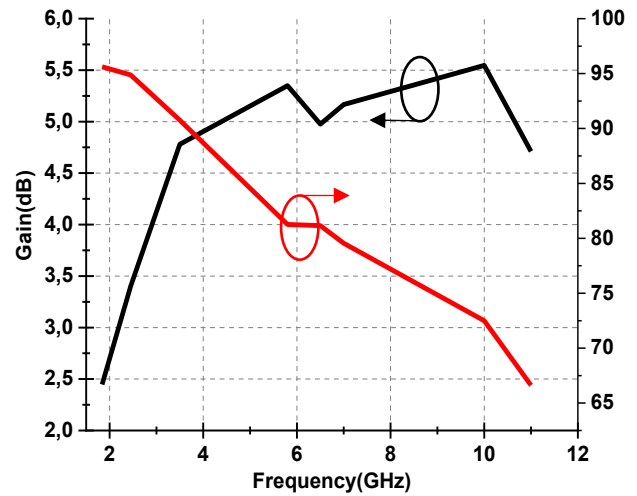


Fig. 4: Simulated Gain and efficiency plot of the proposed antenna

III FABRICATION, MEASUREMENTS, AND DISCUSSION RESULTS

The fabricated prototype is shown in Figure 5. The measured reflection characteristics of the proposed antenna are observed with the help of the VNA NA222A Network Analyser.



Fig. 5: Fabrication prototype of the proposed antenna.

The measured results exhibit good agreement with the simulated results, as shown in Figure 6, demonstrating a similar bandwidth.

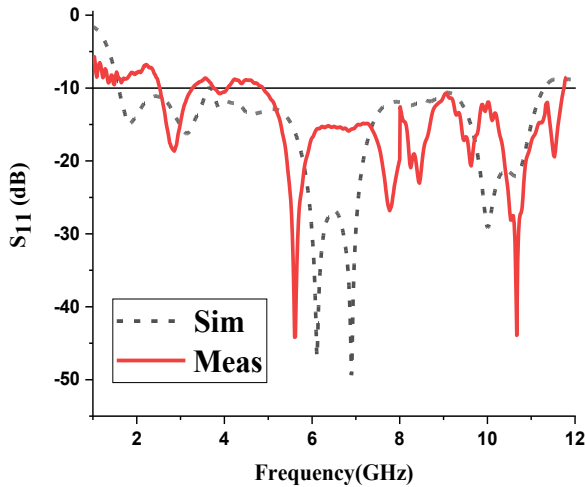


Fig. 6: Simulated and Measured Reflection Characteristics of the proposed antenna in free space.

The proposed prototype for integrating the antenna with the drone is illustrated in Figure 7.

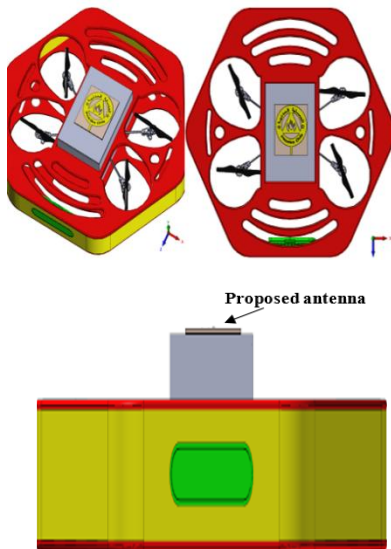


Fig. 7: Positioning of the proposed antenna on the drone.

As depicted, there is excellent compatibility between the antenna and the drone, making our antenna an ideal candidate for this application.

IV. CONCLUSION

This work presented the design, simulation, fabrication, and measurement of an ultra-wideband (UWB) antenna fabricated on FR-4 substrate. The antenna operates effectively from 1.57 GHz to 11.2 GHz, demonstrating good agreement between simulated and measured S11 results. It achieves good gain values (ranging from 2.45 dB to 5.5 dB) across the desired frequency bands while maintaining high efficiency. Furthermore, the antenna's design, inspired by the Algerian Civil Protection Agency emblem, facilitates seamless integration into their equipment such as drones, communication devices, and fire trucks. This integration enables diverse applications in the 1.5 – 11.2 GHz range, including identification, localization, and environmental monitoring in critical situations. Additionally, the antenna's high gain and broadband characteristics make it suitable for independent use in military applications requiring strong signal strength, such as radar and tracking systems, particularly within the C-band. This versatility underscores the antenna's potential as a valuable solution for various military applications.

REFERENCES

- [1] A. Bouaza et al., "Quad-Band Millimeter Wave monopole MIMO Antenna Loaded With a Defected Ground Plane For 5G Applications". In 2024 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), pp. 1-5, Nov. 2024.
- [2] FCC (2002) Federal Communications Commission revision of Part 15 of the commission's rules regarding ultra-wideband transmission systems
- [3] Sami A. Rahman M, Bashir S, "Design of compact tri and quad-band band-pass filters using stub-loaded resonators for wireless applications", SN Appl Sci 1(9):1019, 2019.
- [4] R. George et al., "A low profile reconfigurable antenna for defense aircraft", SN Appl Sci 1(6):608, 2019
- [5] Liang J, Chiau CC, Chen X, Parini CG, "Study of a printed circular disc monopole antenna for UWB systems", IEEE Trans Antennas Propag 53(11):3500–3504, 2005.
- [6] Z. Low, J. Cheong, and C. Law, "Low-cost PCB antenna for UWB applications", IEEE Antennas Wirel Propag Lett 4:237–239, 2005.
- [7] C. W. Ling et al., "Planar binomial curved monopole antennas for ultrawideband communication", IEEE Trans. Antennas Propag 55(9):2622–2624, 2007.
- [8] K-L Wong, S-W Su, C-L Tang, Broadband omnidirectional metal-plate monopole antenna. IEEE Trans Antennas Propag 53(1):581–583, 2005.
- [9] D. Gopi, A. R Vadaboyina, and J.K. Dabbakuti, " DGS-based monopole circular-shaped patch antenna for UWB applications", SN Applied Sciences, 3(2), 198, 2021.
- [10] A. Bouaza et al., "Miniaturization of a Conformal Dual-Band Wearable Textile Monopole Antenna for WBAN and WLAN Applications". International Journal of Communication Systems, 38(7), e70046, 2025.
- [11] A. Bouaza, "High-Performance Flexible Compact Wideband U-Slot CPW-Fed Textile Antenna for WBAN, WLAN, ISM, and WI-FI 6E Applications". International Journal of Communication Systems, 38(4), e70004, 2025.
- [12] A. Bouaza et al. "Performance of a Novel Military Shape Compact Tri-Band Textile Patch Antenna For Short Range Communications Applications". In 2024 International Conference on Advances in Electrical and Communication Technologies (ICAECOT), pp. 1-6, 2024
- [13] Y. Boussaadia, M. Tellache, and F. Amrani, "A Novel High - Gain Rectenna for Wireless Power Transmission (WPT) Applications". International Journal of Communication Systems, 38(9), e70112.
- [14] T. Haq and S. Koziel, "Inverse modeling and optimization of CSRR-based microwave sensors for industrial applications", IEEE Transactions on Microwave Theory and Techniques, t. 70, no 11, p. 4796-4804, 2022.
- [15] Ramesh Garg, "Microstrip antenna design handbook," Boston, Artech House, 2001.
- [16] Constantine Balanis, "Antenna Theory: Analysis and Design," New Delhi, Wiley India, Reprint, 2016.

An MPPT-Based ABC Algorithm Applied to PV Pumping Based BLDC Motor Drive under Partial Shading

Saliha Aoufi, Cherif Larbes, and Aissa Chouder

Abstract— In this paper, we propose a photovoltaic pumping system. Our system is composed of a photovoltaic generator, a boost converter (DC/DC) controlled by Maximum power point tracking (MPPT), an inverter (DC/AC), a motor-pump group (PMBLDC motor, centrifugal pump). The amount of energy produced by PV systems depends on environmental circumstances such as temperature and solar irradiance, the objective of our system is to ensure operation at a maximum power of the photovoltaic system, for this the Artificial Bee method Colony (ABC) was implemented. to evaluate the best and most accurate controller for different weather conditions. The results show that the use of the ABC method is effective and has an excellent performance in terms of response time and can increase the stability of the photovoltaic pumping system. The simulations of the different parts of the system are developed in the MATLAB/Simulink environment.

Keywords— Photovoltaic generator Maximum power point tracking. DC/DC chopper. DC-AC converter. PMBLDC motor. Centrifugal pump.

NOMENCLATURE

ABC	Artificial Bee Colony
DC/DC	DC-DC converter
DC/AC	DC-AC converter
PMBLDC	Permanent magnet brushless direct current
MPPT	Maximum point tracking
PWM	Pulse Width Modulation
EMF	Electromotive force

I. INTRODUCTION

The generation of energy represents a major challenge for the coming years. The energy demands of industrialized nations continue to grow steadily, while developing countries will also require increasing amounts of energy to support their progress and modernization. Currently, a significant portion of global energy production relies on fossil fuels. However, the use of these resources leads to greenhouse gas emissions contributing to, environmental pollution. Moreover, the overconsumption of these non-renewable resources threatens to deplete natural reserves, posing serious risks for future generations [1].

In contrast, renewable energy sources offer a promising alternative, as they are both inexhaustible and can be harnessed in a more environmentally friendly and sustainable manner. A photovoltaic (PV) generator can operate over a wide range of voltage and current values. However, it can only deliver its maximum power at specific voltage and current levels, known as the Maximum Power Point (MPP). The position of this point is highly dependent on environmental conditions such as solar irradiance and temperature, which fluctuate continuously

throughout the day [2]. These variations cause the MPP to shift, making it necessary to implement control strategies that can adapt in real-time without prior knowledge of the MPP location or the cause of its change. This control approach is commonly referred to as Maximum Power Point Tracking (MPPT).

Several MPPT techniques have been developed to optimize the performance of PV systems. Among the most widely used methods are the Perturb and Observe (P&O) algorithm, Fuzzy Logic Control, Artificial Neural Networks, and the Incremental Conductance Method [3]. The primary objective of these algorithms is to continuously track the MPP and ensure efficient energy transfer from the PV generator to the load, regardless of the external environmental changes.

This work aims to ensure optimal performance of the photovoltaic (PV) system under different climatic conditions.[4] To achieve this, we propose an MPPT control technique based on the Artificial Bee Colony (ABC) algorithm for maximum power extraction. The main contributions of this paper are: (i) the design of a new PV water pumping system, and (ii) the direct control of the system using ABC-based MPPT to improve efficiency.

The paper is structured as follows: first part presents the system architecture and modelling, second part describes the ABC algorithm, third part discusses simulation results, and concludes the study.

II. MANUSCRIPT PREPARATION GUIDELINES

Photovoltaic (PV) water pumping system is composed of several key components. It includes a PV generator formed by multiple interconnected solar modules that produce direct current (DC) power. A DC-DC converter is used to regulate the output voltage and frequency according to the available solar energy. Additionally, an inverter converts the DC power into alternating current (AC), while a Pulse Width Modulation (PWM) technique and an MPPT algorithm are implemented to optimize power extraction from the PV modules.

The system also includes a motor-pump assembly consisting of a PMBLDC motor coupled with a centrifugal pump.

Manuscript received April 9, 2026; accepted June 28, 2026.

Saliha Aoufi, Cherif Larbes, and Aissa Chouder are with the Electronics Department, Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: Saliha.Aoufi@g.enp.edu.dz; Cherif.Larbes@g.enp.edu.dz; Aissa.Chouder@g.enp.edu.dz).

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.366

Figure 1 illustrates the overall structure of the proposed system.

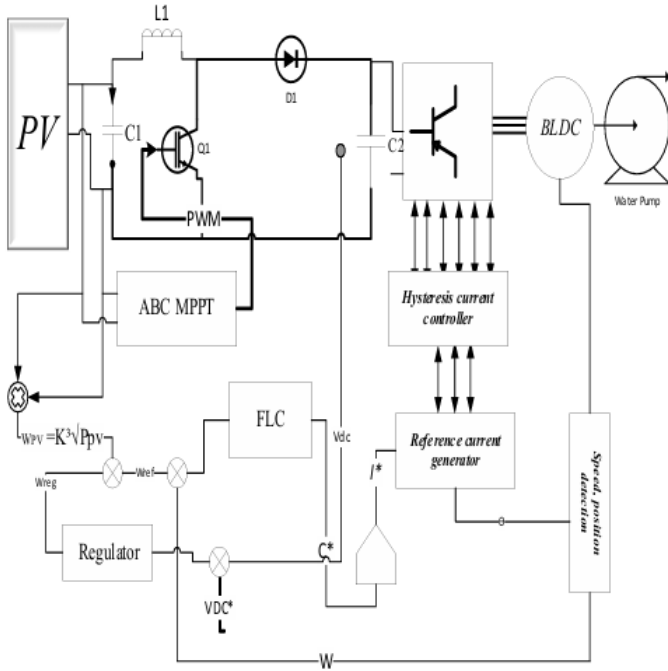


Fig. 1: Bloc Diagram of PV Pumping System

A. Modelling of PV pumping system

In existing literature, several circuit models have been proposed to represent the electrical characteristics of photovoltaic (PV) cells. However, in practical electrical engineering applications, only two models are predominantly used: the Single Diode Model (SDM) and the Double Diode Model (DDM) [5]. Among them, the SDM is most preferred in research due to its simpler structure and lower computational requirements compared to the DDM. The equivalent circuit of the SDM is shown in Figure 2.

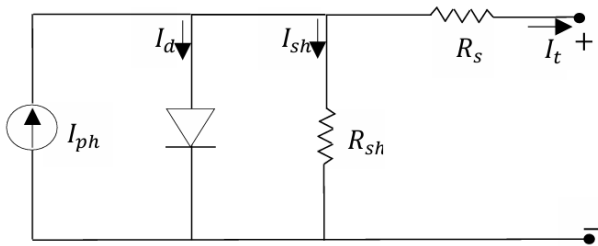


Fig. 2: PV cell equivalent circuit.

This model comprises a photo generated current source in parallel with a diode, a series resistor to denote the ohmic losses related to load current and a shunt resistor to present the leakage current [6]. Thus, in term of Kirchhoff's current law, the PV cell terminal current, it can be expressed by:

$$I_t = I_{ph} - I_d - I_{sh} \quad (1)$$

where I_{ph} denotes the photo generated current, I_d denotes the diode current, and I_{sh} denotes the shunt resistor current, respectively.

Additionally, in term of Shockley equation, I_d is computed by:

$$I_d = I_{sd} \left[\exp \left(\frac{q(V_t + I_t R_s)}{akT} \right) - 1 \right] \quad (2)$$

where I_{sd} is the reverse saturation current of diode, V_t is the cell terminal voltage, R_s is the series resistance, a is the diode factor, k is the Boltzmann constant (1.380×10^{-23} J/K), q is

the electronic charge (1.602×10^{-19} C), and T is the PV cell absolute temperature in Kelvin, respectively.

Moreover, using Kirchhoff's voltage law, I_{sh} is obtained as:

$$I_{sh} = \frac{(V_t + I_t R_s)}{R_{sh}} \quad (3)$$

where R_{sh} is the shunt resistance.

Therefore, by substituting from Equations (2) and (3) into Equation (1), the I-V relationship of the SDM can be rewritten as follows:

$$I_t = I_{ph} - I_{sd} \left[\exp \left(\frac{q(V_t + I_t R_s)}{akT} \right) - 1 \right] - \frac{(V_t + I_t R_s)}{R_{sh}}$$

The specifications of the PV module are given in Table 1

Table 1: PV module parameters

Pmax	Vmp	Imp	Voc	Isc
280	37.01	7.57	43.9	8.14

B. Design of Boost Converter

An energy converter plays a critical role in photovoltaic (PV) systems by serving as an interface between the PV array and the load. Typically, a DC-DC converter is used for this purpose, functioning as a Maximum Power Point Tracker (MPPT). It dynamically adjusts the output voltage of the PV generator to ensure that the system operates at its maximum power point under varying levels of solar irradiance and ambient temperature [7]. This is achieved by controlling the duty cycle of the converter, which regulates the power flow. One of the most used types is the boost converter, also known as a step-up or parallel chopper, which increases the DC voltage from the PV source to a higher level suitable for the load or subsequent stages of the system.

The variation in voltage across the capacitor (output voltage ripple of the Boost converter) is expressed by the following relationship:

$$\Delta V_{C2} = \Delta V_{dc} = \frac{V_{pv}}{R_s C_2 f_p} \frac{D}{1-D}$$

The fictitious load resistance R_s relates the output voltage to the output current in a boost converter. Because this converter maintains power balance (input power equals output power), R_s also reflects the input voltage to input current ratio. To size the output capacitor C_2 for the nominal maximum power point of a photovoltaic system (1000 W/m^2 at 25°C), R_s is defined based on specific input voltage and current values.

$$R_s = \frac{V_{c2}}{I_2} = \frac{V_{dc}}{I_2} = \frac{V_{pv}}{I_2} = \frac{V_{mp}}{I_2}$$

The output voltage ripple of a Boost converter depends on system parameters and the duty cycle D . As D increases, the amplitude of the ripple also increases. To size the capacitor C , it is advisable to consider a worst-case scenario by using the maximum value of D , which may lead to over-sizing C if the operation occurs at a lower D .

Additionally, selecting a high switching frequency f is

preferable to reduce the required capacitance of the capacitor. Assuming a linear variation of the current in the inductor, this variation or ripple of the current is given by:

$$\Delta I_e = \frac{U_{ei}}{L_e f} D$$

The current ripple is directly proportional to the duty cycle D.

To size the inductance L, the maximum D should be considered to keep the ripple within limits. Choosing a high switching frequency f_p also helps reduce the required inductance and limit the ripple. The variation in voltage across the input capacitor C1 is expressed by:

$$\Delta V_{c1} = \frac{\Delta I_1}{8f_p C_1}; \quad C_1 = \frac{\Delta I_1}{8f_p \Delta_{c1}}$$

The amplitude of the input voltage variation (ΔV_{in}) is proportional to the current ripple in the input inductor (ΔI_1) and inversely proportional to the capacitor value C1 and the switching frequency (f_p). To limit (ΔV_{in}), one can increase the input capacitor C1 or the switching frequency (f_p).

C. Modelling of the BLDC Motor Drive System

Theoretically, PMBLDC (Permanent Magnet Brushless DC) and AC machines share operational principles with conventional DC machines; however, they replace the mechanical collector with an electronic commutator. This transition enhances efficiency and reduces maintenance needs. In the context of conventional DC generators connected in parallel to a DC network, BLDC motors exhibit a superior torque-to-size ratio, making them particularly valuable in applications where space and weight constraints are critical, such as aerospace, robotics, and portable electronics [8].

Moreover, BLDC motors are designed to generate trapezoidal waveform electromotive forces (EMFs) while being supplied with rectangular stator currents. This configuration not only helps achieve a consistent torque output but also improves overall efficiency and performance. Additionally, the precise control of these motors allows for rapid response times and smooth operation, making them suitable for high-performance applications. The electronic commutation also enables advanced control strategies, such as field-oriented control, which further optimize their performance under dynamic conditions [9].

Table 2: Six step switching sequence for commutation

Rotor position	H1	H2	H3	Switch Closed	Phase current		
					A	B	C
0-60	1	0	0	S1 S4	+	-	0
60-120	1	1	0	S1 S6	+	0	-
120-180	0	1	0	S3 S6	0	+	-
180-240	0	1	1	S3 S2	-	+	0
240-300	0	0	1	S5 S2	-	0	+
300-360	1	0	1	S5 S4	0	-	+

The BLDC motor's uniqueness stems from its trapezoidal

electromotive force, resulting in non-sinusoidal mutual inductance between the stator and rotor. The parameter, φ_0 , is the magnitude of the magnetic flux created by the permanent magnet and the electrical angle. Thus, it yields

$$\vec{V} = [R]\vec{I} + \frac{d}{dt}\{[L]\vec{I}\} + \frac{d}{dt}\vec{\varphi}_M$$

where V and I are vectors representing voltage, current respectively for each phase, φ_M is the vector of the magnetic flux created by the permanent magnet, [R] et [L] are the resistance and inductance matrices of the machine.

For a symmetrical three-phase winding and a balanced system, the vector of the voltages at the terminals of the three phases is given by:

$$\begin{pmatrix} v_a \\ v_b \\ v_c \end{pmatrix} = \begin{pmatrix} R & 0 & 0 \\ 0 & R & 0 \\ 0 & 0 & R \end{pmatrix} \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} + \frac{d}{dt} \begin{pmatrix} L_{aa} & L_{ab} & L_{ac} \\ L_{ba} & L_{bb} & L_{bc} \\ L_{ca} & L_{cb} & L_{cc} \end{pmatrix} \begin{pmatrix} i_a \\ i_b \\ i_c \end{pmatrix} + \begin{pmatrix} e_a \\ e_b \\ e_c \end{pmatrix}$$

with e represents the electromotive force.

The mechanical model of the BLDC motor is given by:

$$J \frac{d}{dt} w_r = C_e - f w_r - C_r$$

where w_r is the angular velocity, in rad/s, C_e is the electromagnetic torque N/m, C_r is the load torque N/m and f is the coefficient of friction on N.(m/ rad).s⁻¹.

To generate trapezoidal currents in the machine windings, a 120° control strategy with a single hysteresis regulator is employed. This method uses a hysteresis band around the reference current to control transistor signals based on the measured current's intersection with the band limits. It is cost-effective, requiring a 6-sector position sensor and one or two current sensors. The operation is divided into six 60° sectors, activating only two machine phases at a time, except during switching [10]. The magnitude of the phase current (I^*) is determined from the reference torque: $I^* = C^* / K$. Depending on the position of the rotor, the reference current generator block produces three phase currents (IA^* , IB^* , IC^*) by assigning the reference current values (I^* , $-I^*$, and zero).

D. Modelling of the Pump

The Q-H curve represents the relationship between lift height and discharge, with manufacturers providing maximum and minimum curves based on wheel diameter. A centrifugal pump operates with three key parameters: height, flow, and speed. [28] Using the motor's rotational speed as a parameter, the "PELEIDER-PETERMAN" model is expressed as follows:

$$H_n = a_1 w^2 + a_2 w q + a_3 q^2$$

in which, a_1 , a_2 and a_3 are geometric parameters characterizing the pump. It should be noted that the other expression to consider is that of the power in Watts absorbed by a pump, i.e. the power required for its mechanical drive, which is expressed by:

$$P = 9.81 \frac{QH}{n_p}$$

The pump efficiency is defined as the ratio of the hydraulic power imparted by the pump to the fluid to the shaft mechanical power [11].

III. MPPT CONTROL TECHNIQUES FOR PV SYSTEM

A. Artificial Bee Colony

The ABC algorithm is a population-based, iterative optimization technique modelled after the intelligent foraging behaviour of honeybees. It consists of three main types of bees [12]:

- **Employed Bees:** Search for food sources (possible solutions) and share the information with onlooker bees.
- **Onlooker Bees:** Select food sources based on the quality (fitness) shared by employed bees and then further exploit the most promising ones.
- **Scout Bees:** Explore the environment randomly to find new food sources (diversify the search to avoid local maxima).

The flowchart of the ABC algorithm is given in Figure 3.

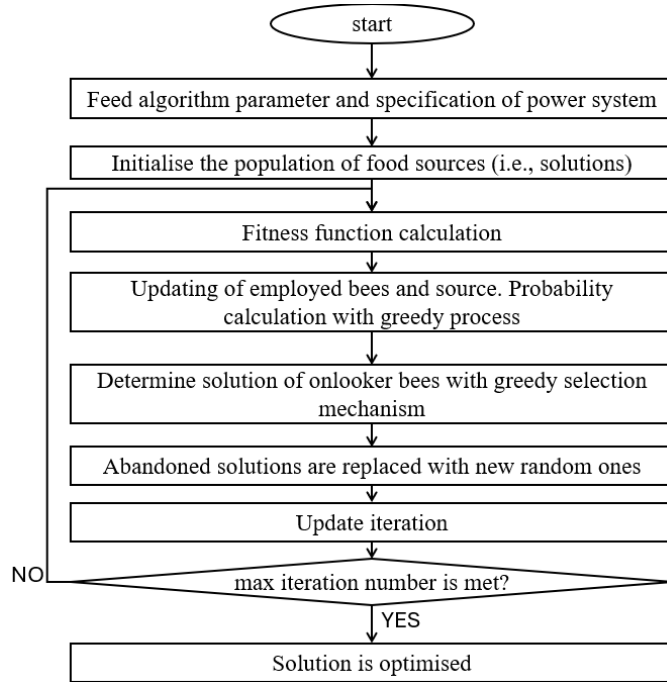


Fig. 3: Flowchart of ABC algorithm.

As can be seen in this figure, the process is as follows:

1. Initialization

Parameters Setup: Define the number of solutions SN (duty cycles), maximum cycle number MCN, and sampling time Ts, ensuring Ts exceeds the power converter's settling time.

Initial Population: Generate a randomly distributed initial population of solutions (duty cycles) within the range [0, 1].

2. Employed Bees Phase

Solution Update: Each employed bee modifies its current solution by exploring a neighbouring solution, aiming to improve its fitness.

Evaluation: Calculate the fitness of the new solution. •

Greedy Selection: If the new solution is better, update the current solution; otherwise, retain the existing one.

$$d_i = d_{min} + \text{rand}[0, 1][d_{max} - d_{min}]$$

3. Onlooker Bees Phase

Probability Calculation: Determine the probability of selecting each solution based on its fitness.

Solution Selection: Onlooker bees probabilistically select solutions and explore their neighbourhoods.

Evaluation and Selection: Evaluate the new solutions and apply greedy selection as in the employed bee's phase.

4. Scout Bees Phase

Exploration: Scout bees randomly generate new solutions to replace those that are stagnated or less fit.

Evaluation: Assess the fitness of the new solutions.

Update: Replace the old solutions with the new ones if they are better as below

$$\text{new}(d_i) = d_i + \Phi_i(d_i - d_j)$$

5. Repeat

Repeat the above phases until the maximum cycle number MCN is reached or the power value remains unchanged within a specified number of cycles.

B. Effect of Irregular Solar Radiation

The PV generator, which consists of five modules connected in series, is subjected to various forms of solar irradiation, including uniform and non-uniform (partial shading), as shown in Table 4, in order to examine the behaviour of the chosen PV pumping system. Three modes of sunlight are in fact taken into consideration: two distinct non-uniform modes of sunlight and a uniform mode of sunlight (case 2). (Participatory appeal of Cases 1 and 3. 4).

Table 3: Solar irradiation modes (W/m2) for each module of the PV system

CASE 1	CASE 2	CASE 3	CASE 4
1000	800	1000	800
1000	800	1000	800
1000	800	1000	600
1000	700	1000	600
900	800	1000	1000
800	600	800	800
800	600	1000	1000

To evaluate the performance of the PV system under different irradiance conditions, the power-voltage (P-V) and current-voltage (I-V) characteristics of the PV array were analysed. Figure 4 presents the obtained characteristics for both uniform irradiance and partial shading conditions, highlighting the location of the maximum power points (MPPs) under each operating scenario. The figure clearly shows the existence of a single MPP under uniform irradiance and multiple local MPPs under partial shading conditions.

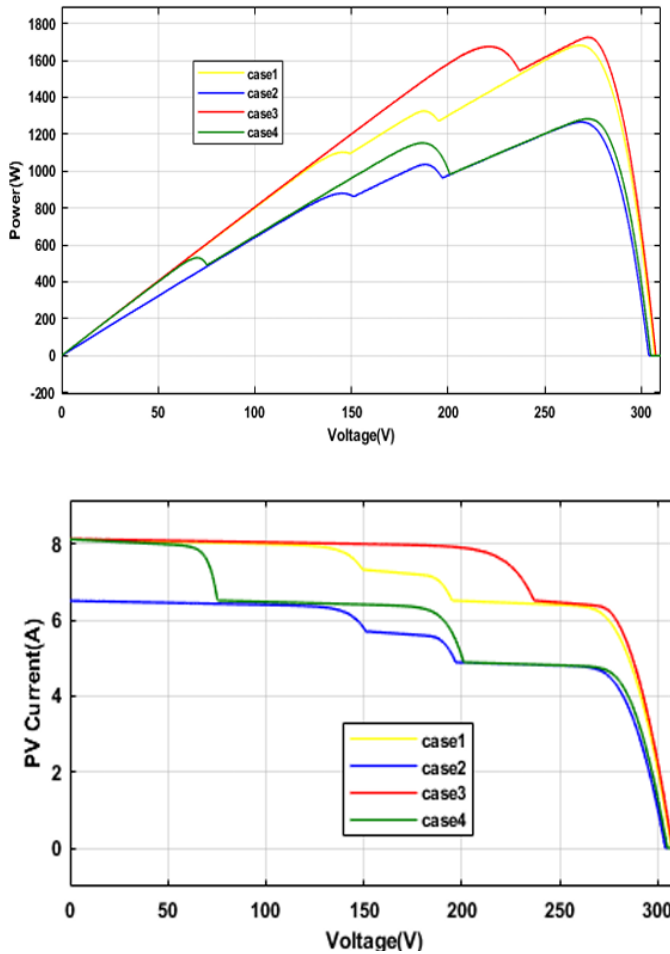


Fig. 4: P-V and I-V characteristics of the PV array under different irradiance conditions.

The uniform irradiance condition of Case 1 is characterized by the presence of a single maximum power point (MPP), with a power output of 1682 W corresponding to a PV generator voltage of 251 V and a current of 6.7 A (Figure 4).

The partial shading condition of Case 2 is characterized by the presence of three local maximum power points and one global maximum power point. The global maximum power achieved under this condition is 1265 W, corresponding to a PV generator voltage of 268 V and a current of 4.8 A (Figure 4).

Based on these observations, controlling the output voltage of the PV generator is essential to ensure operation at the maximum power point under varying temperature and irradiance conditions. This improves both the power conversion efficiency of the PV generator and the overall system performance. To achieve this objective, the MPPT-ABC algorithm is employed, which has demonstrated fast convergence and reliable performance when applied to the PV generator-Boost converter system.

To assess the effectiveness of the proposed ABC-MPPT algorithm, the dynamic responses of the PV power, PV current, and PV voltage were monitored under different irradiance conditions. The corresponding simulation results are presented in Figure 5.

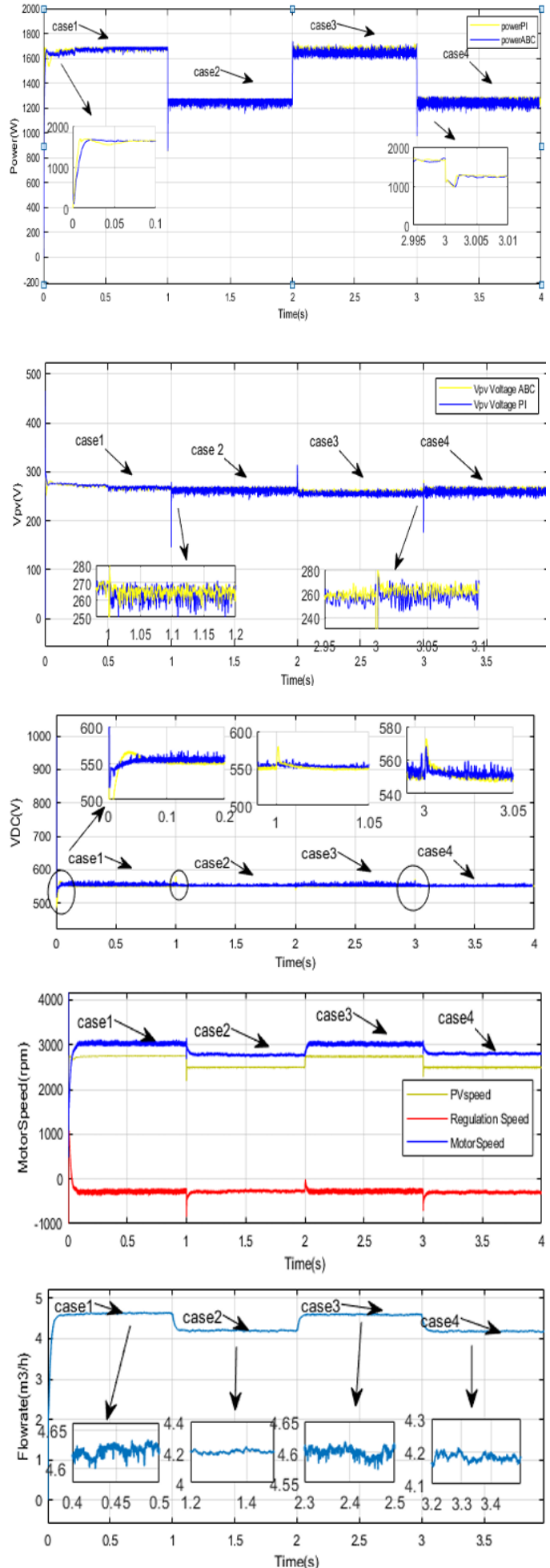


Fig. 5: Dynamic response of the ABC-MPPT algorithm: (a) PV output power, (b) PV array current, and (c) PV array voltage.

Under the first irradiance condition (Case 1), the ABC-MPPT algorithm achieved a steady-state PV power of approximately 1678 W after a response time of 0.02 s (Figure 5a). This corresponds to an efficiency of 99.71% relative to the theoretical maximum power of 1682 W shown in Figure 4.

The voltage across the PV array was measured at 260 V, while the current reached 6.3 A (Figures 5b and 5c). These values are very close to the reference maximum power point values of 261 V and 6.5 A, respectively (Figure 4). In this case, the duty cycle stabilized at 0.4033. The corresponding load voltage and current are illustrated in Figures 5c and 5b, respectively.

Under the second irradiance condition (Case 2), the steady-state PV power reached approximately 1260 W after a response time of 0.01 s (Figure 5a). The ABC-MPPT algorithm maintained an efficiency of 99.6% relative to the theoretical maximum power of 1265 W (Figure 4). The PV array voltage was 261 V, while the current was 4.7 A (Figures 5c and 5b).

Similar results were obtained for the partial shading conditions of Cases 3 and 4. The ABC-MPPT algorithm enabled the PV array to reach 99.83% of its actual global maximum power. The corresponding output powers were approximately 1721 W and 1283 W for Cases 3 and 4, respectively (Figure 5a). The system response time was approximately 0.01 s. The PV array voltage was measured at 257 V, while the current reached 3.3 A (Figures 5c and 5b), which are very close to the reference maximum power point values of 251 V and 3.39 A.

These results demonstrate that, under different irradiance conditions, the ABC-MPPT algorithm is capable of accurately tracking the global maximum power point with high efficiency, fast response, and excellent stability.

C. Efficiency Estimation of BLDC Motor–Pump System

A very good efficiency is obtained for the proposed water pumping system. Table 7 and Figure 7. show the efficiency estimation of the BLDC motor-pump system and its graphical representation respectively, subjected to the variation in atmospheric condition, we present the variations in the water flow rate of the pump (Q) and power with irradiation variation.

As the solar insolation level changes, various indices of the BLDC motor-pump system, such as back EMF (ea), stator current (isa), speed (W), electromagnetic torque, and load torque, also vary in response to the solar insolation, as illustrated in les Figures precedent. Two key observations can be drawn from the simulation results. First, the stator current (isa) is controlled during startup, allowing it to gradually reach its steady-state value, resulting in a soft start for the BLDC motor. Second, the electromagnetic torque generated by the BLDC motor consistently matches the torque required to drive the pump, demonstrating the system's stable operation under varying solar insolation conditions, regardless of whether fluctuations.

In the solar pumping system without energy storage, all the electricity produced by the photovoltaic generator (Ppv) is directly used by the pump motor (Pm). It is essential to maintain a constant balance between Ppv and Pm to prevent any instability in the system. This requires controlling the rotational speed of the pump motor, which must correlate with the mechanical power (Pm) of the motor to align with the photovoltaic power (Ppv). Thus, the reference speed is calculated based on Ppv. In parallel, a regulation speed is introduced, with an opposite sign to Wpv, to obtain the total reference speed This regulation speed allows for the control of

the charge and discharge power of the bus capacitor, ensuring that the voltage remains close to its reference value.

Table 4: Parameters of BOOST and BLDC motor

PARAMETER	VALUE	PARAMETER	VALUE
Boost		BLDC	
L (mH)	5.7	Stator phase resistance (Ω)	2.875
C1 (μ F)	1.728	Stator phase inductance (mH)	8.5
C2 (μ F)	57.51	Flux linkage (V.s)	0.175
f (kHz)	50	Inertia (kg.m^2)	8e-4
		Viscous damping (N.ms)	e-3
		Pole pair	4

A. CONCLUSION

This study focused on modelling and analysing the components of a photovoltaic (PV) water pumping system, which includes an BLDC motor pump powered by a PV generator through a BOOST converter and an inverter. An analysis was conducted on Maximum Power Point Tracking (MPPT) technique: the ABC (Artificial Bee Colony. Among these, the algorithm demonstrated superior performance, maintaining consistent operation at maximum power with rapid and accurate tracking even under changing weather conditions. Its robustness was verified under various solar radiation scenarios. The results revealed that the ABC method outperformed the other two in terms of response time, resistance to climate variation, and water flow efficiency. Moreover, it enabled effective operation even at very low solar irradiance levels.

Overall, the proposed system significantly enhances the efficiency of PV-based water pumping. Future work will explore the integration of hybrid optimization algorithms for broader photovoltaic applications.

REFERENCES

- [1] M. B. Jebli and M. Kahia, "The interdependence between CO2 emissions, economic growth, renewable and non-renewable energies, and service development: Evidence from 65 countries," *Climate Change*, vol. 162, no. 2, pp. 193–212, 2020.
- [2] J. Shi, W. Zhang, Y. Zhang, F. Xue, and T. Yang, "MPPT for PV systems based on a dormant PSO algorithm," *Electric Power Systems Research*, vol. 123, pp. 100–107, 2015.
- [3] Z. Zhao, R. Cheng, B. Yan, J. Zhang, Z.-H. Zhang, M. Zhang, and L. L. Lai, "A dynamic particle MPPT method for photovoltaic systems under partial shading conditions," *Energy Conversion and Management*, vol. 220, 2020.
- [4] G. Li, Y. Jin, M. W. Akram, X. Chen, and J. Jia, "Application of bio-inspired algorithms in maximum power point tracking for PV systems under partial shading conditions—A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 840–873, 2018.
- [5] A. A. S. Mohamed, A. Berzoy, and O. Mohammed, "Design and hardware implementation of FL-MPPT control of PV systems based on GA and small-signal analysis," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 279–290, Jan. 2017.
- [6] M. V. da Rocha, L. P. Sampaio, and S. A. O. da Silva, "Comparative analysis of MPPT algorithms based on Bat algorithm for PV systems under partial shading conditions," *Sustainable Energy Technologies and Assessments*, vol. 40, Art. no. 100761, 2020.
- [7] A. Ahmed and Z. Salam, "An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency," *Applied Energy*, vol. 150, pp. 97–108, 2015.
- [8] R. Kumar and B. Singh, "BLDC motor driven solar PV array fed water pumping system employing Zeta converter," *IEEE Transactions on Industry Applications*, doi: 10.1109/TIA.2016.2522943.

- [9] H. B. Hansen, C. S. K. Kallesøe, and J. D. Bendtsen, "A hybrid model of a brushless DC motor," in Proc. 16th IEEE International Conference on Control Applications, 2007.
- [10] A. Terki, A. Moussi, A. Betka, and N. Terki, "An improved efficiency fuzzy logic control of PMBLDC for PV pumping systems," Mathematical Modelling, vol. 36, pp. 934–944, 2012.
- [11] S. Meunier, M. Heinrich, L. Quéval, J. A. Chemic, L. Vido, et al., "A validated model of a photovoltaic water pumping system for off-grid rural communities," Applied Energy, vol. 241, pp. 580–591, 2019.
- [12] X. Bi and Y. Wang, "Artificial bee colony algorithm with fast convergence," Systems Engineering and Electronics, vol. 33, no. 12, pp. 2755–2761, 2011.

FPGA Implementation for Fast and Accurate Medical Image Segmentation

Fatma Zohra Hamidatou-Hamadi, Mohamed Lamine Hamidatou and Latifa Hamami-Mitiche

Abstract—This work presents a comprehensive study and hardware implementation of two advanced medical image segmentation methods: Gradient Vector Flow and Distance Regularized Level Set Evolution. Thesis approaches are widely used in image processing due to their ability to accurately extract complex contours, particularly in concave and noisy regions. However, their high algorithmic complexity leads to significance computational cost, limiting their use in real-time applications. To address these limitations, implementation on a Field-Programmable Gate Array FPGA is proposed, leveraging hardware parallelism to enhance performance. A simulation phase was conducted using MATLAB on real medical images (brain and breast MRI scans). The results demonstrate that the GVF method is effective in handling concave boundaries, while the DRLSE method exhibits greater robustness to topological changes and multiple objects. The hardware implementation was carried out using Vitis HLS and Vivado on a PYNQ-Z2 platform. The results show strong agreement with MATLAB simulations, with negative discrepancies due to numerical precision, along with a significant improvement in processing time.

Keywords—DRLSE, FPGA, GVF, Level Set.

NOMENCLATURE

DRLSE	Distance Regularized Level Set Evolution
FPGA	Field-Programmable Gate Array
GVF	Gradient Vector Flow
LUT	Look-Up Table
MRI	real medical images

I. INTRODUCTION

Medical image segmentation is a fundamental step in many clinical applications, including diagnosis, tumor detection, and therapeutic monitoring [1]. It involves extracting structures of interest from medical images, such as MRIs or CT scans. Among the most widely used methods, active contours and Level Set methods hold an important place due to their ability to model complex shapes. However, these methods have certain limitations. Traditional active contours are highly dependent on initialization and have difficulty converging in concave areas [2]. To improve this performance, the GVF method was introduced, allowing better capture of concavities [3]. Level Set methods, while efficient at handling topological changes, require frequent resetting, which increases computation time. The DRLSE model has been proposed to eliminate this constraint by incorporating a regularization term [4].

Despite their performance, these methods remain computationally expensive. Therefore, FPGA implementation

offers a promising solution for accelerating their execution.

II. ACTIVE CONTOURS AND GVF

1. ACTIVE CONTOURS

Active contours are evolving curves defined by a parametric function (Figure 1). Their evolution is controlled by minimizing a total energy composed of an internal energy and an external energy [5].

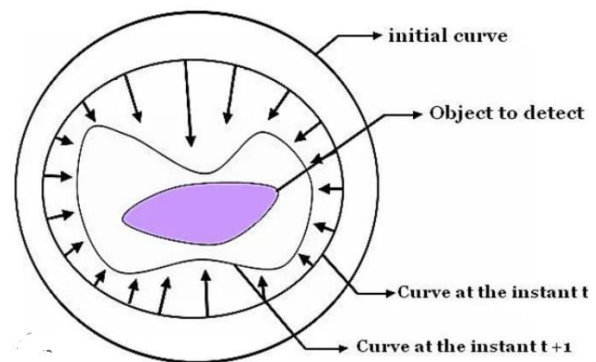


Fig. 1: Representation of the evolution of the active contour.

The outline E defined by a representation parametric [5], the E assimilated has a curve C represented according to equation (1):

$$C = \left\{ \begin{matrix} V(s, t) = \\ [x(s, t), y(s, t)] \end{matrix} \right\}, \forall (s, t), [0, 1] \times [0, \infty] \quad (1)$$

with :

- s : abscissa curvilinear the long of outline.
- t : evolution temporal of there curve In the image.
- $V(s, t)$: position of a point of there curve C has A instant t .

Manuscript received April 9, 2026; accepted June 28, 2026.

Fatma Zohra Hamidatou-Hamadi, Mohamed Lamine Hamidatou and Latifa Hamami-Mitiche are with the Ecole Nationale Polytechnique, Algiers, ALGERIA.

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.367

The energy functional (equation 2) associated with the active contour consists of two terms:

$$E_{snake}(v) = E_{interne}(v) + E_{externe}(v) \quad (2)$$

➤ Energy internal.

The internal energy of an active contour regulates the shape of the curve by ensuring its continuity and controlling its curvature (equation 3).

$$E_{int}(v) = \frac{1}{2} \int_0^1 (\alpha(s) \cdot |v'(s)|^2 + \beta(s) |v''(s)|^2) ds \quad (3)$$

With :

$$|v'(s)|^2 = \left(\frac{dx(s)}{ds}\right)^2 + \left(\frac{dy(s)}{ds}\right)^2$$

$v'(s)$: derivative first of $v(s)$ by report has s .

v'' : derivative second of $v(s)$ by report has s .

α : elasticity of outline.

β : rigidity of outline.

The internal energy of an active contour can be broken down into two fundamental components: the continuity term and the curvature term.

➤ Energy external

According to Kass , Witkin and Terzopoulos , the external energy of an active contour (equation 4) consists of two principal components [2] :

$$E_{ext}(v) = E_{image}(v) + E_{contrainte}(v) \quad (4)$$

Image energy,

Image energy guides the curve C toward object boundaries by relying on image features such as gradients and contours (equation 5):

$$E_{image}(v) = - \int_0^1 \lambda(s) \left\| \nabla(I(v(s))) \right\|^2 ds \quad (5)$$

where $\nabla I(v(s))$ is the image gradient and $\lambda(s)$ is a weighting factor.

High-gradient regions attract the curve toward relevant edges. Internal energy enforces smoothness, while external energy drives the contour to image boundaries. However, this approach has limitations, particularly in handling concave shapes.

2. GRADIENT VECTOR FLOW.

To improve the basic snake method, C. Xu and J.L Prince developed the GVF snake method [3] introducing a new external force called Gradient Vector Flow (GVF) (equation 6):

$$F_{ext} = v = [u(x, y), v(x, y)] \quad (6)$$

The calculation of the GVF field takes place in two stages:

A. Calculation of the Contour Map (Edge) Map)

B.

C.

The contour map $f(x, y)$ of image $I(x, y)$ is determined by the derivative of the image. It is often expressed as follows (equation 7):

D.

$$f(x, y) = |\nabla(G_\sigma(x, y) * I(x, y))|^2 \quad (7)$$

where ∇ is the gradient and G_σ is a Gaussian filter with standard deviation σ .

This contour map generates vectors perpendicular to the contours, with high amplitudes near the contours and low amplitudes in homogeneous areas. In these homogeneous areas, a diffusion process is applied to orient the vectors also towards concave areas. This approach improves the convergence of active contours in areas of high concavity and around acute angles.

B. Calculation of flow of vector gradient:

The flux of the gradient vector is defined as a vector field

$V(x, y) = (u(x, y), v(x, y))$ that minimizes the following energy functional (equation 8):

$$\varepsilon = \iint (\mu_1 (u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 \cdot |V - \nabla f|^2) dx dy \quad (8)$$

with :

$$u_x = \frac{\partial u}{\partial x}, u_y = \frac{\partial u}{\partial y}, v_x = \frac{\partial v}{\partial x}, v_y = \frac{\partial v}{\partial y}$$

in which, μ is a setting of regularization between the first and the second term of the integral.

This energy functional allows the GVF vector field to be optimized to achieve better convergence of active contours in complex areas, particularly those with concavities and acute angles.

3. DESCRIPTION OF THE GVF ALGORITHM

To concisely illustrate the various steps involved in the GVF segmentation method, the flowchart presented in Figure 2 provides an overview of the algorithm workflow, beginning with image preprocessing and concluding with the generation of the final segmented image.

Prior to applying the GVF method for image segmentation, an essential preprocessing stage is performed to enhance the quality of the extracted contours. Anisotropic filtering is particularly effective in this context, as it reduces image noise while preserving important structural features and edge information within the medical image.

Following preprocessing, the contour map is obtained by applying a Laplacian-of-Gaussian (LoG) filter, which enhances edge detection and improves the identification of relevant object boundaries. The external force components, (u) and (v), are calculated from the image gradients derived using Sobel operators. These components are subsequently regularized through Laplacian smoothing to stabilize their values during the iterative GVF diffusion process.

During each iteration, the contour evolution is updated by computing the total energy, (E), and comparing it against a predefined convergence tolerance. To ensure a uniform distribution of contour points along the snake, an interpolation procedure is incorporated to maintain an appropriate average spacing between consecutive points. This process requires defining a minimum distance, (dmin), and a maximum distance, (dmax), which establish the allowable range for the distance between neighboring contour points. The GVF method improves the convergence and robustness of active contours, particularly when segmenting objects with concave boundaries. This is achieved by extending the influence region of the image gradient forces, enabling the snake to effectively detect and adapt to complex object geometries.

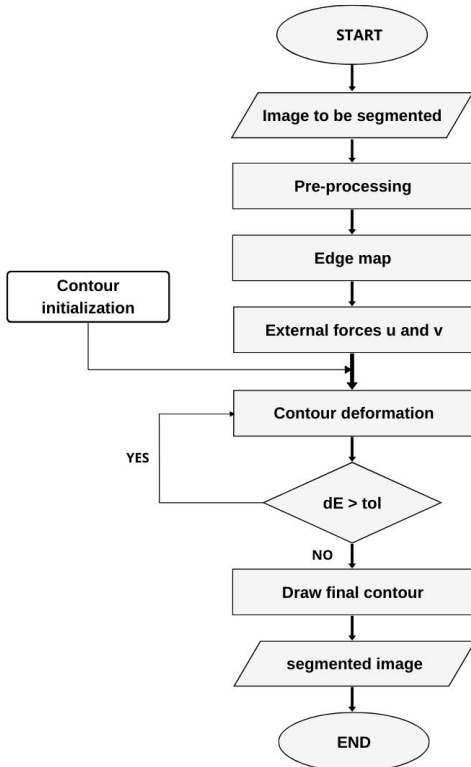


Fig 2: Flowchart of The GVF method

III. Regularized Level-Set Evolution Distance Method DRLSE

The DRLSE method is an evolution of Level Set methods. It introduces a regularization term that allows the distance function to be maintained without requiring resetting. This new approach uses an energy function composed of three terms, equation (9):

$$E(\phi) = \mu P(\phi) + \lambda L_g(\phi) + \alpha A_g(\phi) \quad (9)$$

Regularization term that allows the function to maintain its distance function form throughout the evolution.

$$P(\phi) = \int \frac{1}{2} (|\nabla \phi| - 1)^2 dx dy$$

Length term: The energy function $L_g(\phi)$ Allows the active contour to position itself close to the contour of the object of interest.

$$L_g(\phi) = \int g \delta_\epsilon |\nabla \phi| dx dy$$

δ_ϵ : Dirac function, defined as follows equation (10):

$$\delta_\epsilon(x) = \begin{cases} 0 & |x| > \epsilon \\ \frac{1}{2\epsilon} \left[1 + \cos\left(\frac{\pi x}{\epsilon}\right) \right] & |x| \leq \epsilon \end{cases} \quad (10)$$

With ϵ : regularization parameter.

Surface term : the energy function $A_g(\phi)$ Allows the speed of evolution of the active contour to be accelerated when the latter is far from the object of interest for a faster convergence.

$$A_g(\phi) = - \int g H(-\phi) dx dy.$$

Where $H(-\phi)$ is the Heaviside function,

The DRLSE model [4] incorporates a distance regularization term to stabilize the contour evolution function, thereby improving segmentation speed and eliminating the need to reset the function. The evolution equation then becomes (equation 11):

$$\frac{\partial \phi}{\partial t} = \mu \text{div}(d_p(|\nabla \phi|) \nabla \phi) + \lambda \delta_\epsilon(\phi) \text{div}\left(g \frac{\nabla \phi}{|\nabla \phi|}\right) + \alpha g \delta_\epsilon(\phi) \quad (11)$$

μ : Distance regularization parameter.

λ : Contour length regularization parameter.

α : Surface repulsion parameter.

d_p is defined by equation 12:

$$d_p(s) \cong \frac{p'(s)}{s} \quad (12)$$

Where p is a potential function used for distance regularization. For better distance regularization, the double-well potential function p_2 is used, which is expressed as follows:

$$P_2 = \begin{cases} \frac{1}{2\pi} (1 - \cos(2\pi s)), & \text{si } s \leq 1 \\ \frac{1}{2} (s - 1)^2, & \text{si } s > 1 \end{cases}$$

1. DRLSE ALGORITHM.

The organizational chart of the DRLSE method, describing its operation, is presented in Figure 3.

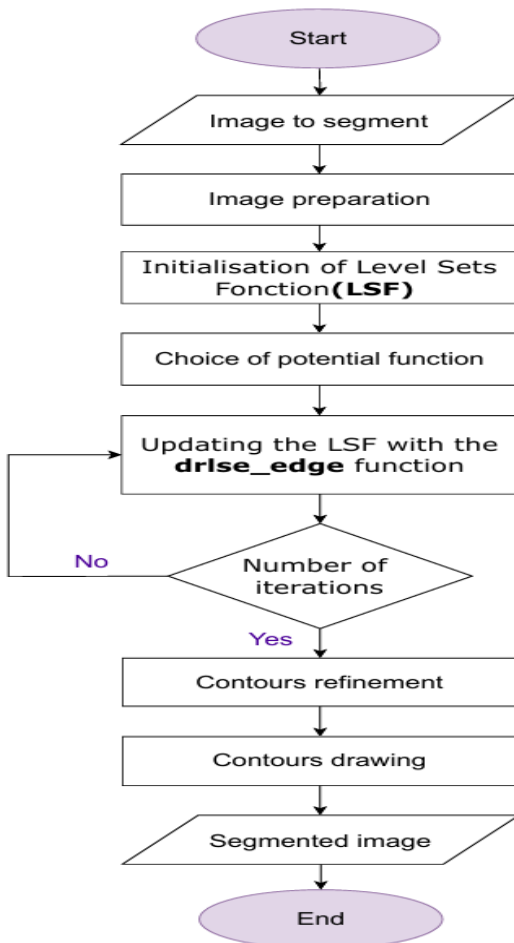


Fig 3: Flowchart of the DRLSE Method.

The proposed DRLSE algorithm starts with an image preparation stage, where the input image undergoes preprocessing to enhance its suitability for segmentation. During this stage, the user also specifies the parameters required for the segmentation process. Next, the level set function (LSF) is initialized as a binary function, with a

rectangular region defining the initial contour that will evolve toward the object boundaries. If required, multiple contours can be initialized simultaneously by specifying several initial regions.

The contour evolution process is then performed through a main iterative loop executed over a predefined number of outer iterations. During each iteration, the `drlse_edge` function updates the level set function (ϕ) according to the Distance Regularized Level Set Evolution (DRLSE) model.

The DRLSE formulation combines several energy terms to guide the contour evolution. The weighted length encourages smooth and regular contours by minimizing the contour length. The distance regularization term maintains the stability and regularity of the level set function by preserving its signed-distance property. In addition, the weighted area term drives the contour toward the target object by promoting the segmentation of interior regions while suppressing irrelevant boundaries. Together, these energy terms enable accurate and stable object segmentation.

The choice of the potential function is a key factor in the performance of level set-based segmentation methods. A single-well potential function is generally suitable for segmenting homogeneous regions, where object boundaries are relatively simple and a clear separation between regions is required. On the other hand, a double-well potential function is more effective for images containing complex structures, as it can simultaneously capture edge information and homogeneous interior regions. By providing a richer energy formulation, the double-well potential improves the stability of the evolving contour and enables it to accurately follow both sharp transitions and smooth image features.

Level set-based image segmentation methods, particularly those based on Distance Regularized Level Set Evolution (DRLSE), offer effective control of contour regularization while maintaining numerical stability throughout the evolution process. In addition, they provide accurate boundary detection and robust edge localization, making them well suited for segmenting complex images with irregular shapes and varying intensity distributions.

IV. IMPLEMENTATION ON FPGA.

Our implementation is based on the PYNQ-Z2 FPGA board, a development platform using the SoC Xilinx Zynq-7000. This board integrates both a Processing System (PS) and Programmable Logic (PL), offering an integral solution for the development of embedded hardware and software [8].

1. IMPLEMENTATION STEPS

Figure 4 shows a synoptic diagram of the steps we followed during the implementation.

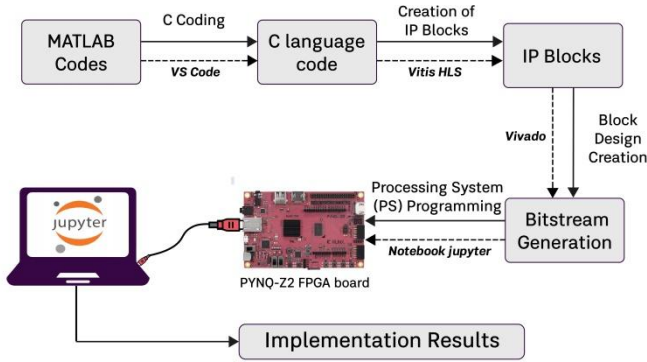


Fig. 4: Simplified synoptic diagram of the HLS approach

First, we write our program in C using Visual Studio Code (VS Code), then we run a simulation to ensure the program functions correctly by comparing the results with those generated in MATLAB. Before generating VHDL code, we perform exhaustive tests and simulations on our C code in Vitis HLS. This allows us to detect and correct potential errors earlier in the development process [6]. Vitis HLS also allows us to generate an IP Block. Using Vivado, we create a design block for the main system, which will be deployed on an FPGA and will include the previously generated IPs [7], [8]. Once the Block Design approved, we created A HDL Wrapper Who converts the plan of architecture in program HDL (Hardware Description Language). We synthesize the project, then we carry out the implementation and generate the binary file " Bitstream ".

The FPGA card is programmed using the generated Bitstream file and the system's proper functioning is verified.

2. GVF ALGORITHM ON FPGA

In this section, we detail the implementation steps of the GVF algorithm on an FPGA circuit (Figure 5). After developing a complete version of the algorithm, we will explain each step of the process in detail, as illustrated in the block diagram above.

We divided the first part of the GVF algorithm into two distinct blocks: one block for calculating the contour map f and another for the Sobel filter as well as the calculation of the variables bn , $c1$, and $c2$. The outputs of the first block are stored in DDR memory and then duplicated for retrieval by the second block. This process is also applied to the other intermediate variables for the two remaining blocks, which are dedicated to calculating the external force fields and the contour deformation.

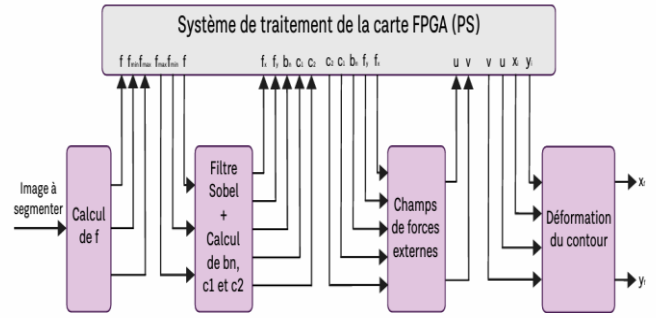


Fig. 5: Complete diagram of the GVF algorithm

Each block corresponds to a top function that we develop in C language and simulate on VS Code, and therefore to an IP block that we create on Vitis HLS.

The "contour map" function takes a filtered image as input to calculate the horizontal and vertical gradients (f_x and f_y), the gradient amplitude (bn), and the curvature characteristics ($c1$ and $c2$) necessary for contour detection. The "external force fields" function uses these values to generate the external force fields u and v . Finally, the "contour deformation" function receives the external fields u and v , along with the coordinates of the initial contour (x_i and y_i), and calculates the coordinates of the final contour (x_f and y_f).

Each block is designed to perform a specific function optimally, enabling robust and efficient implementation of the algorithm in an FPGA environment. Each step is rigorously tested, and the overall system operation is verified to achieve performance and accuracy targets.

3. DRLSE ALGORITHM ON FPGA.

The first stage of the implementation is divided into three distinct blocks (see Figure 6): convolution, calculation of (g) , and gradient computation. This subdivision enables more efficient management of hardware resources by storing the output data from each block in DDR memory and subsequently duplicating it for use by the following processing stages. Each block consists of a top-level function developed in the C language and simulated using Visual Studio Code, as well as an IP block generated using Vitis HLS for hardware implementation.

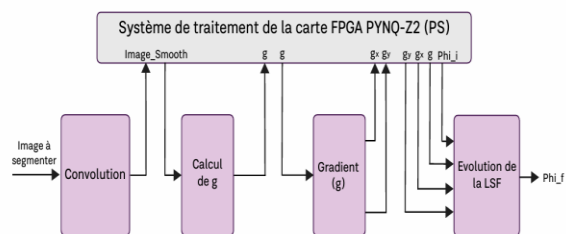


Fig. 6: Complete diagram of the DRLSE algorithm

The "LSF Evolution" function takes the initial distance function ϕ_i , g , g_x , and g_y and outputs ϕ_f . At each iteration, ϕ_f is calculated using an evolution equation that combines these data. The image contours (g) act as an external force that pulls the boundary of ϕ towards the detected contours in the image, allowing the LSF to accurately follow the contours. The gradients (g_x) and (g_y) are used to adjust the boundary to follow changes in image intensity and texture, ensuring precise and detailed segmentation.

The resulting block design provides a modular structure for the efficient implementation of the DRLSE algorithm on FPGAs. Each block is designed to optimally execute a specific function, enabling a robust and high-performance implementation of the algorithm within an FPGA environment. Each step is rigorously tested, and the overall system operation is verified to achieve performance and accuracy targets.

V. RESULTS AND INTERPRETATIONS.

The segmentation results obtained using the two algorithms simulated in MATLAB for medical images are presented. Additionally, the FPGA implementation is described. The efficiency of the proposed method in segmenting these images is evaluated by adjusting the algorithm parameters and analyzing their impact on the segmentation performance.

For the FPGA design, a block-based approach is adopted, in which predefined modules are utilized. Each block represents a specific function, such as processing units or memory controllers, and the blocks are interconnected through standardized interfaces to facilitate system integration and evaluate their influence on the segmentation results.

1. MATLAB SIMULATION

Before segmentation, preprocessing is applied to improve image quality. This includes:

- anisotropic filtering
- noise reduction

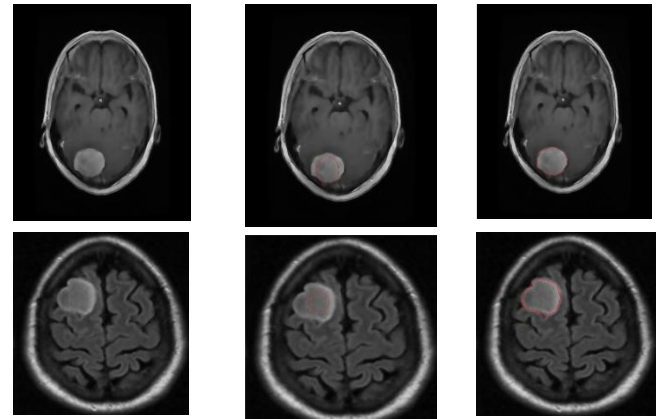
The GVF and DRLSE methods were tested on medical images (brain MRI, breast).

A. Results of Segmentation by the GVF Method .

➤ *Results of MRI images of a brain.*

After optimizing the parameters and initialization of the active contour for the GVF method, we are now applying this segmentation technique to medical images, including brain and breast MRIs.

Figure 7 illustrates the results obtained by applying the GVF method to MRI images of a brain with the parameters, $\mu = 0.15$, $\alpha = 0.1$, $\beta = 0.06$, $\text{tol} = 0.004$, $\text{itmax} = 300$.



(a) Picture initial (b) Outline initial (c) Result

Fig. 7: Segmentation with GVF On MRI of a brain

➤ *Image results MRI of a breast.*

Breast MRI segmentation using the GVF method has encountered difficulties. Breast MRIs, characterized by intensity variations and complex textures, complicate the application of GVF. Noise and similar anatomical structures disrupt the external force field, leading to incorrect convergence of the active contours. Consequently, the segmented contours do not accurately correspond to the actual boundaries of the structures of interest.

B. Results of segmentation using the DRLSE method.

➤ *Results of MRI images of a brain.*

To segment these images, we empirically adjusted the parameters α , λ and μ to find the optimal combination that allows efficient segmentation (figure 8).

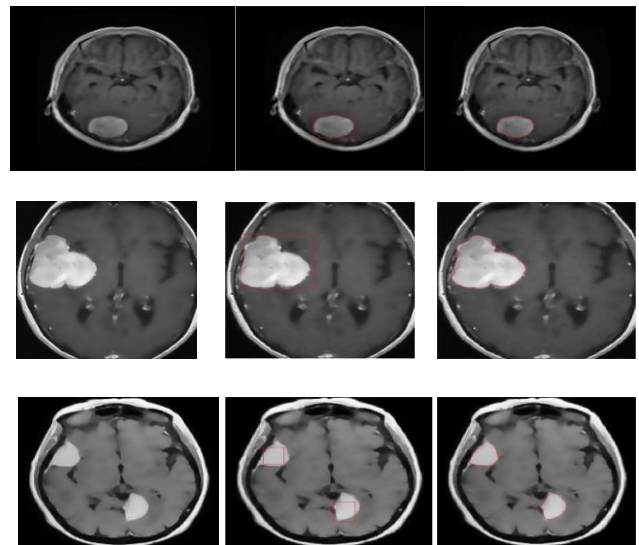
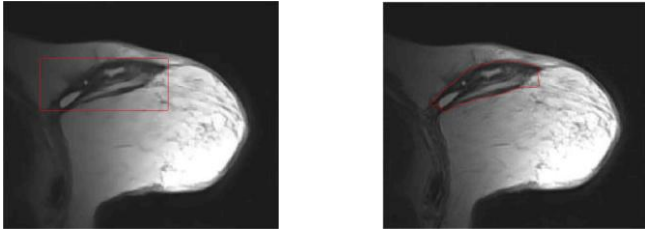


Fig. 8: Segmentation on brain MRIs $\alpha = 0.5$, $\lambda = 6$ and $\mu = 0.02$.

➤ Results on MRI images of a breast

After several parameter adjustments, applying the DRLSE algorithm to this new medical image yielded a less precise result in terms of segmenting the structures of interest. This observation led us to consider other approaches to improve the segmentation quality. Figure 9 shows the segmentation result with the parameters: $\alpha = 1.5$, $\lambda = 5$, and $\mu = 0.02$.



(a) Outline initial

(b) Picture segmented

Fig. 9: Segmentation result with the DRLSE algorithm on breast MRI

2. FPGA IMPLEMENTATION

A. Implementation using the GVF method on FPGA

Each block represents a specific function, such as processing units or memory controllers, and the blocks are interconnected through standardized interfaces to enable communication between them. This modular approach facilitates efficient utilization of FPGA resources, thereby simplifying the overall FPGA design process. Figure 10 presents the final block-based design of the proposed architecture for the complete implementation of the GVF algorithm [9].

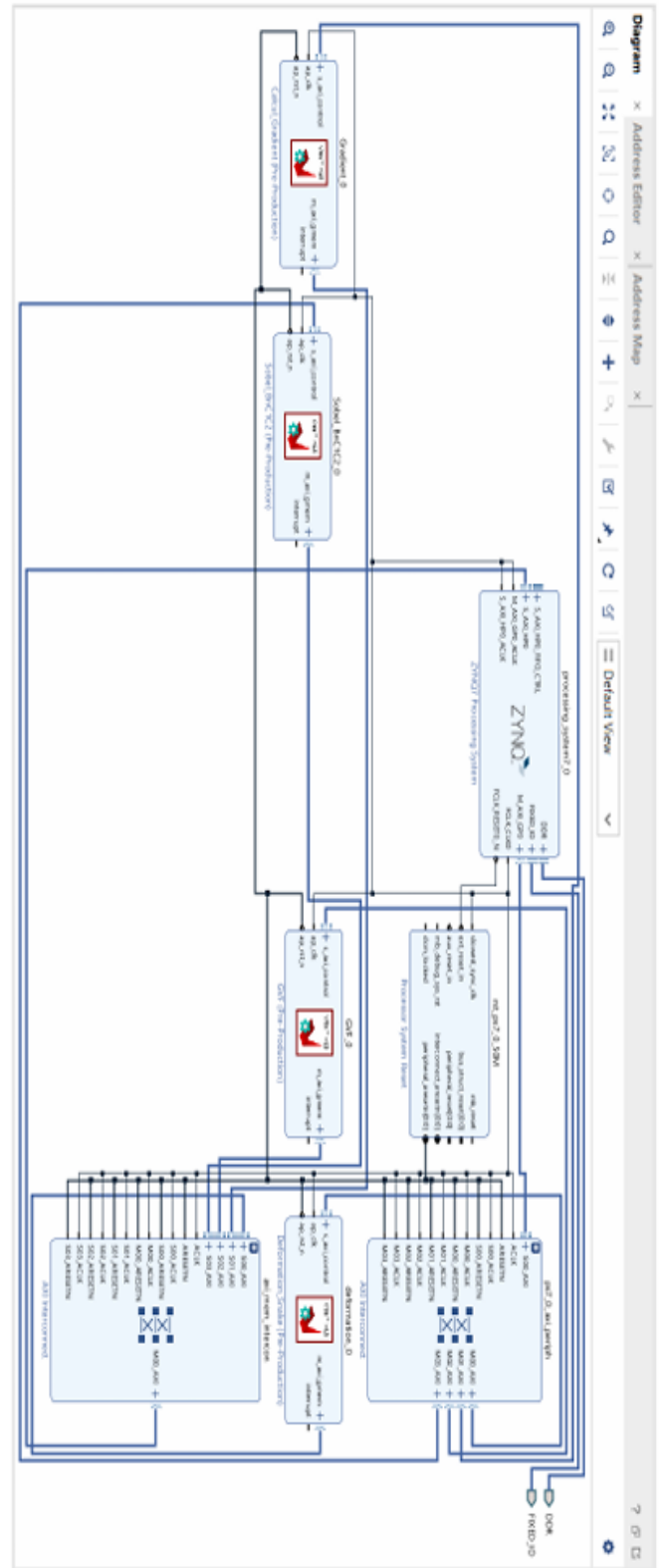
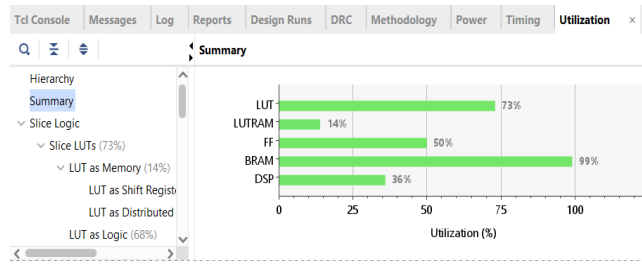
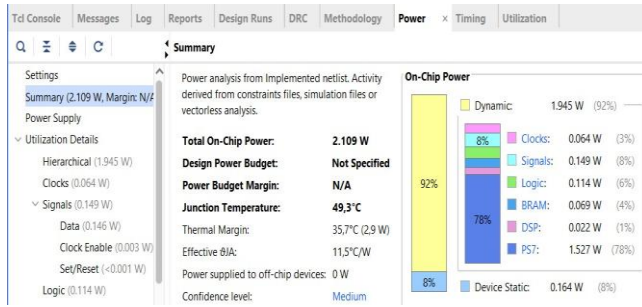


Fig. 10: Block design for the complete implementation of the active contour algorithm using the GVF method

Figure 11 presents the utilization and power ratios of complete architecture



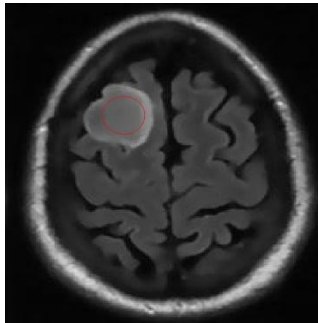
Usage report



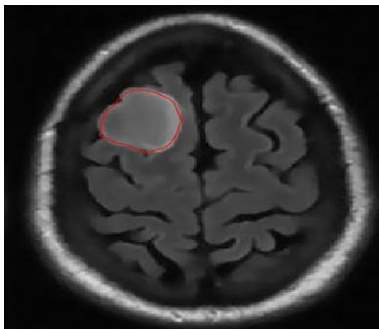
Report on power

Fig. 11: Implementation of The GVF algorithm on FPGA

By maintaining identical parameter settings and the same initial contour, the results obtained from the MATLAB simulation and the FPGA implementation are visually equivalent. It can be observed that segmentation accuracy is preserved, with the FPGA-based results showing a similar visual appearance to those obtained in MATLAB (Figure 12). The execution time of the complete GVF architecture is measured at 12.786 s.



Initial contour



segmented image on FPGA

Fig. 12: Result of the implementation of GVF on FPGA

The slight differences observed stem primarily from numerical precision. MATLAB uses double-precision (64-bit) floating-point numbers, while the FPGA implementation relies on single-precision (32-bit). This reduced precision leads to small deviations during calculations, particularly in repetitive operations such as gradient calculations (f , f_x , f_y), where rounding errors can accumulate. Furthermore, optimizations applied during synthesis with Vitis HLS can also introduce minor variations.

B. Implementation of the DRLSE

The final block-based design of the proposed architecture for the complete implementation of the DRLSE algorithm is illustrated in Figure 13 [10].

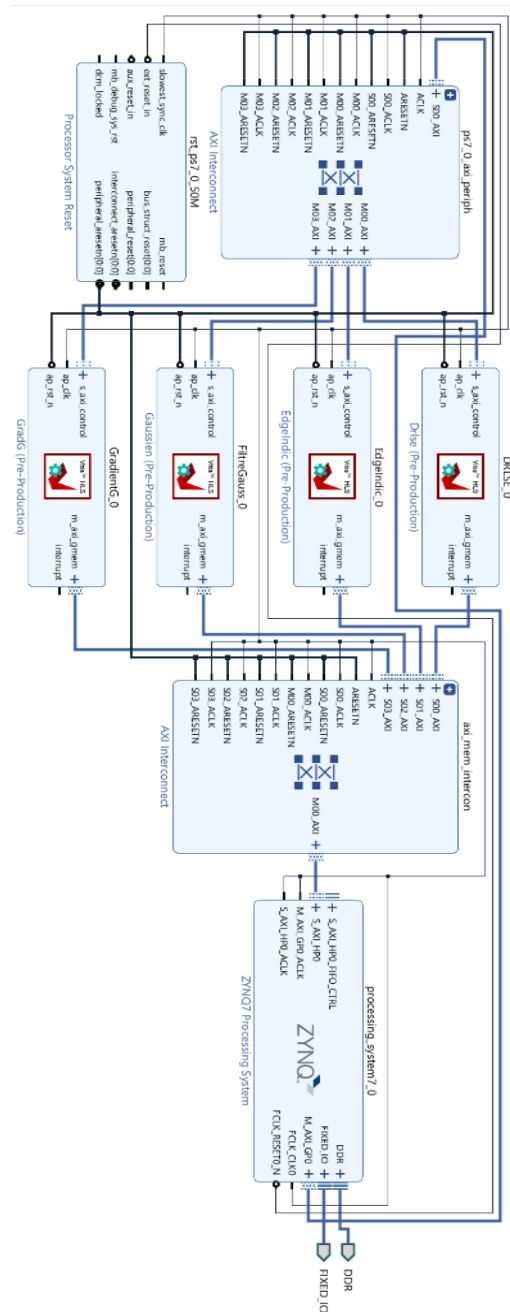


Fig. 13: Design by blocks for the implementation of the algorithm DRLSE

Each block represents a specific function, such as processing units or memory controllers, and the blocks are interconnected through standardized interfaces to enable communication between them. By maintaining the same parameter configuration and initial contour used in the MATLAB simulation, the segmentation results obtained from the FPGA implementation of the DRLSE method are visually equivalent to those obtained previously. The corresponding segmentation result is presented in Figure 14.

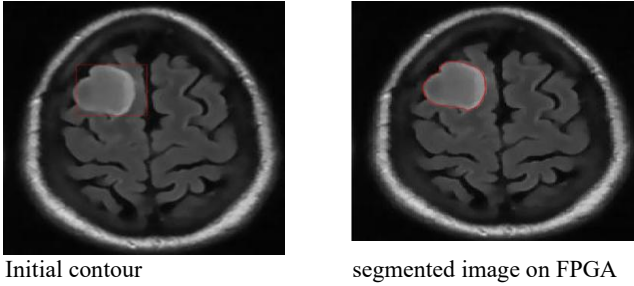


Fig. 14: Result of the implementation of their method DRLSE

The FPGA implementation of the DRLSE algorithm demonstrated satisfactory segmentation accuracy, with a rigorous visual comparison performed against the results obtained from the MATLAB simulation. Figure 15 presents the resource utilization and power consumption analysis of the DRLSE algorithm implemented on the FPGA. It highlights the main hardware resources, including lookup tables (LUTs), block RAMs (BRAMs), and digital signal processing slices (DSPs), as well as the estimated power consumption.

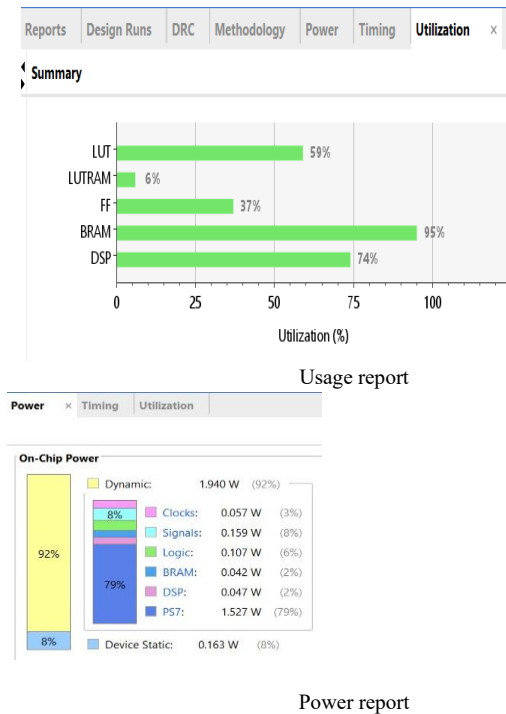


Fig. 15: Implementation of the algorithm DRLSE on FPGA

VI. COMPARISON OF SEGMENTATION RESULTS: GVF VS DRLSE ON FPGA.

The two algorithms, GVF and DRLSE, were implemented on an FPGA and tested under similar conditions. The results show good visual accuracy for both approaches, with detected contours close to those obtained using MATLAB. GVF is efficient for concave and complex shapes, with moderate FPGA resource consumption. DRLSE provides sharper and more precise contours, in particular

For geometric structures, but with greater use of hardware resources (DSP, BRAM). In terms of performance:

- The execution time is slightly faster for GVF compared to DRLSE.
- The numerical differences between the FPGA and MATLAB results are minimal and do not impact visual quality.

GVF is lighter, while DRLSE offers finer segmentation, at the cost of higher hardware complexity.

To achieve these results, we optimized our calculations by grouping them into a single execution. This required the use of buffers for matrix updates, thus increasing the use of computing resources and extending the overall execution time, measured at 105.8523 seconds. A distinctive feature of our approach was the frequent subdivision of a block into sub-blocks, favoring the use of DDR memory rather than the FPGA's BRAM. This strategy aimed to maximize the efficiency of available memory, even though data accesses memory is generally slower than the logic part of the FPGA.

VII. CONCLUSION

This work made it possible to compare two advanced methods of medical image segmentation and to demonstrate the value of their implementation on FPGA.

The GVF is effective for concave contours, while the DRLSE is more robust for complex structures.

The FPGA implementation demonstrates remarkable efficiency in utilizing hardware resources, fully leveraging the inherent parallelism of this architecture. It enables a significant reduction in execution time while maintaining high accuracy. FPGAs prove particularly well-suited to complex medical image segmentation algorithms, as they facilitate the optimization, reuse, and adaptation of modules according to processing needs. They thus offer an excellent balance between performance, accuracy, and flexibility, while paving the way for more scalable and high-performance hardware architectures for real-time image processing.

REFERENCES

- [1] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," arXiv preprint arXiv:1811.02629, 2018.
- [2] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988, doi: 10.1007/BF00133570.
- [3] C. Xu and J. L. Prince, "Gradient vector flow: A new external force for snakes," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Juan, PR, USA, 1997, pp. 66–71, doi: 10.1109/CVPR.1997.609407.
- [4] Y. Chen, G. Chen, Y. Wang, and N. Dey, "A distance regularized level-set evolution model based MRI dataset segmentation of the brain's caudate nucleus," *IEEE Access*, vol. 7, pp. 133093–133104, 2019, doi: 10.1109/ACCESS.2019.2937964.
- [5] J.-J. Rousselle, "Active contours: A segmentation method. Application to medical imaging," Ph.D. dissertation, 2003.
- [6] A. Jarrah and S. Amri, "Optimized FPGA-based implementation of brain tumor detection by combining K-means and gray wolf optimization algorithms," *Traitement du Signal*, vol. 39, no. 6, pp. 1879–1891, 2022, doi: 10.18280/ts.390601.
- [7] AMD Xilinx, *Vitis High-Level Synthesis User Guide (UG1399)*, Version 2021.2, 2021. [Online]. Available: <https://docs.amd.com/r/2021.2-English/ug1399-vitis-hls>
- [8] AMD Xilinx, *Vivado Design Suite User Guide (UG973)*, Version 2021.1, 2021. [Online]. Available: <https://docs.amd.com>
- [9] M. L. Hamidatou, F. Z. Hamadi, L. Hamami, B. Bouzid, S. A. Belkacem, and F. Z. Allam, "Hardware implementation of the GVF approach for deformable contour detection on FPGA," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 30239–30245, Dec. 2025.
- [10] F. Z. Hamadi, M. L. Hamidatou, L. Hamami-Mitiche, S. A. Belkacem, and B. Bouzid, "MATLAB simulation and FPGA implementation of the DRLSE segmentation algorithm," *Traitement du Signal*, vol. 42, no. 2, pp. 719–727, Apr. 2025, doi: 10.18280/ts.420210..

computer science and physics. He is also a Research Master at the LMCS laboratory of the Higher National School of Computer Science (ESI) in Algiers. After obtaining a graduate diploma in Electronics from USTHB, Algiers, he earned a master's degree in electronic communications and Signal Processing from the University of Blida, Algeria, followed by a Doctorate (Doctorat en Sciences) in the field of image and speech processing from the National Polytechnic School (ENP) of Algiers. Several final-year student projects have been carried out under his supervision in the areas of image and speech processing.

Latifa Hamami-Mitiche (M.Sc. 1988, Ph.D. 2002) is a full Professor in the Department of Electronics at Ecole Nationale Polytechnique of Algiers, Algeria. She is author more than 150 scientific papers. Her research interests include Medical Image Processing, ISAR Image Processing, Optical Character Recognition, Biometrics and Embedded System, FPGA and ASIC Design, System on Chip.

Fatma Zohra Hamidatou-Hamadi is an Assistant Professor at the National School of Agronomy. She teaches physics courses at ENSA. She is the author of several international publications. Her research focuses on image processing, particularly medical images. Her objectives are multiple, including the implementation of intelligent embedded systems dedicated to medical diagnosis support on FPGA-based embedded circuits.

Hamidatou Mohamed Lamine is currently a Senior Lecturer (Maître de Conférences, Classe A) at the National School of Agronomy (ENSA) in Algiers, where he teaches courses in

SCHO-Optimized Fractional-Order Control Strategy for Enhanced Alternator Voltage Stability

Elouahab Bouguenna, Samir Ladaci, Bilal Taghezouit, Amine Berouaken, and Walid Merrouche

Abstract— This paper proposes a novel metaheuristic optimization framework grounded in the mathematical properties of hyperbolic trigonometric functions, introducing the Sinh–Cosh Optimizer (SCHO) as a bio-inspired algorithm that leverages the inherent gradient-like dynamics of sinh and cosh operators to achieve a principled and adaptive equilibrium between global exploration and local exploitation across complex search landscapes. The proposed framework is applied to tune an incommensurate Fractional-Order PID (FOPID) controller for Automatic Voltage Regulator (AVR) systems, where nonlinear dynamics and parametric uncertainties pose critical control challenges. The incommensurate non-integer differintegral structure extends the controller's degrees of freedom, enhancing its capacity to handle memory-dependent system behavior. Comparative simulations against conventional PID and standard FOPID benchmarks demonstrate a 32.115% overshoot reduction and a 33.335% ITSE improvement, with sustained robustness under load disturbances and parametric perturbations. The SCHO–FOPID strategy offers a scalable and mathematically rigorous intelligent control solution for uncertain nonlinear power systems, with direct applicability to smart grid and renewable energy contexts.

Keywords— sinh-cosh optimizer, fractional calculus, alternator voltage stability, robustness.

I. INTRODUCTION

In modern power generation infrastructures, synchronous generators constitute the backbone of energy conversion from renewable sources to regulated electrical output. Nevertheless, terminal voltage deviations — induced by abrupt load transitions, mechanical torque oscillations, and electromagnetic disturbances — pose severe threats to the integrity of voltage-sensitive loads, including microprocessor-based systems and communication networks, where even marginal excursions can trigger irreversible failures [1], [2]. Automatic voltage regulators (AVRs) are therefore integrated into synchronous generator units as closed-loop excitation control mechanisms, continuously adjusting field current to confine output voltage within prescribed operational bounds.

Classical AVR implementations have long relied on proportional–integral–derivative (PID) controllers; however, the inherent rigidity of their integer-order structure renders them inadequate in the presence of system nonlinearities, unmodeled dynamics, and time-varying parametric uncertainties. Fractional calculus — which generalizes the order of differentiation and integration to the continuum of real numbers — has emerged as a mathematically rich and practically powerful paradigm for advanced control synthesis. Fractional-order PID (FOPID) controllers augment the classical three-parameter structure with two additional degrees of freedom, namely the integral order λ and the

derivative order μ , yielding superior transient shaping, tighter reference tracking, and markedly improved disturbance rejection relative to their integer-order counterparts [3], [4]. From an analytical perspective, optimal FOPID parameterization necessitates resolving ill-conditioned fractional differential systems — a class of problems intrinsically amenable to population-based stochastic search [5]. Accordingly, metaheuristic optimization has established itself as the dominant paradigm for non-integer controller tuning, with representative contributions spanning chaos-augmented atom search strategies [6] and various hybrid swarm-based architectures [7]–[11] demonstrating consistent gains in closed-loop convergence and error minimization. Crucially, the No Free Lunch theorems (NFLTs) [12]–[15] formally establish that no optimizer holds universal dominance across arbitrary problem landscapes, thereby sustaining the imperative for algorithm diversity and continued methodological innovation. Preserving signal smoothness, solution correctness, and scalable performance under unmodeled dynamic conditions nonetheless remains an active and unresolved challenge.

Motivated by these considerations, this work investigates the Sinh–Cosh Optimizer (SCHO) [16], a recently introduced metaheuristic whose search mechanics are governed by the complementary analytical properties of hyperbolic trigonometric functions: the cosh component drives broad global exploration across the feasible space, while the sinh component orchestrates fine-grained local exploitation, collectively yielding a self-regulating and well-balanced optimization dynamic. Within the AVR control problem, SCHO is leveraged to identify the optimal parameter set of an incommensurate FOPID controller, targeting stable closed-loop behavior, accelerated convergence, and minimized steady-state tracking error under realistic disturbance and uncertainty scenarios.

The principal contributions of this study are threefold: (i) establishing the aptitude of SCHO for high-dimensional fractional-order tuning problems; (ii) formulating a stability-informed, constraint-aware optimization framework for

Manuscript received June 24, 2026; accepted July 2, 2026.

Elouahab Bouguenna and Samir Ladaci are with the Ecole Nationale Polytechnique, Algiers, ALGERIA. Bilal Taghezouit, Amine Berouaken, and Walid Merrouche are with CDER, Algeria.

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.393

FOPID synthesis; and (iii) demonstrating quantifiable dynamic performance gains under structured AVR parametric variations.

II. MATHEMATICAL MODEL OF AVR

An AVR constitutes a closed-loop excitation control architecture whose fundamental objective is to maintain the generator terminal voltage at a prescribed setpoint, irrespective of load fluctuations and varying operating conditions. By continuously confining the output voltage within allowable bounds, the AVR safeguards the reliability, power quality, and dynamic stability of the electrical supply delivered to downstream loads [17]. A representative configuration of an AVR-integrated power generation unit is illustrated in Fig. 1.

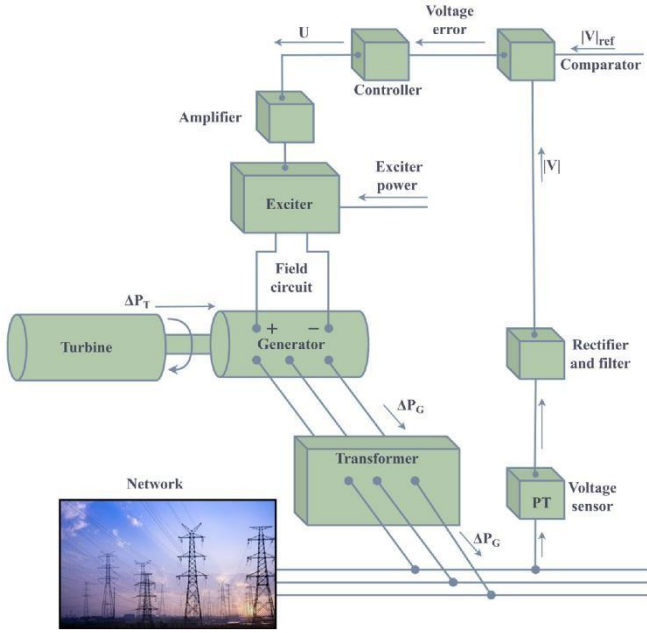


Fig. 1: Block diagram representation of the AVR loop.

In its basic operating principle, the AVR acquires a measurement of the terminal voltage, evaluates its deviation from a reference command, and processes the resulting error through an amplification stage to synthesize the appropriate control action. The generated signal is subsequently forwarded to the excitation subsystem, which modulates the generator field current and, consequently, the air-gap magnetic flux, steering the output voltage toward the target setpoint.

The detailed closed-loop structure of the regulated AVR system is depicted in Fig. 2, in which the synchronous generator acts as the primary controlled plant. A dedicated voltage transducer continuously acquires the output voltage and feeds it back to the summation node, where it is contrasted against the reference command. The resulting error signal quantifies the instantaneous discrepancy between the measured terminal voltage and its desired value. The controller interprets this error to compute the corrective excitation input in accordance with the governing control law, which acts upon the field current to drive the terminal voltage into convergence with the reference trajectory.

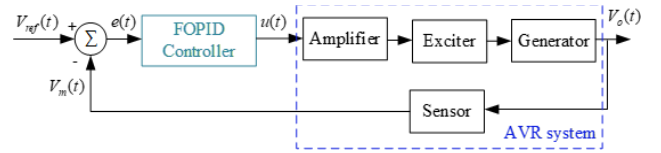


Fig. 2: Block diagram of a closed-loop AVR with feedback regulation.

Within this feedback architecture, the control design objectives transcend simple steady-state regulation. They encompass closed-loop stability preservation, robustness against parametric uncertainties and unmodeled system dynamics, bounded and smooth control effort, and minimization of residual tracking error. For analytical tractability, the voltage dynamics of an AVR-equipped synchronous generator are commonly represented by a first-order transfer function approximation [18]. Despite its simplicity, this reduced-order model adequately captures the dominant input–output voltage response characteristics and provides a practical and well-established foundation for controller synthesis and closed-loop stability assessment.

For reference, Table I compiles the transfer function models most widely adopted in the AVR literature, along with their associated parameter ranges, serving as a unified modeling basis for the comparative analysis conducted throughout this study.

Table. I

OPERATING LIMITS OF AVR SYSTEM PARAMETERS			
Components	Transfer function	Gain	Time Constants (s)
Amplifier	$H_A(s) = \frac{G_a}{1 + s\tau_a}$	$10 \leq G_a \leq 400$	$0.02 \leq \tau_a \leq 0.1$
Exciter	$H_E(s) = \frac{G_e}{1 + s\tau_e}$	$1 \leq G_e \leq 10$	$0.4 \leq \tau_e \leq 1$
Generator	$H_G(s) = \frac{G_g}{1 + s\tau_g}$	$0.7 \leq G_g \leq 1$	$1 \leq \tau_g \leq 2$
Sensor	$H_S(s) = \frac{G_s}{1 + s\tau_s}$	$1 \leq G_s \leq 2$	$0.001 \leq \tau_s \leq 0.06$

In which τ_a , τ_e , τ_g , and τ_s represent the amplifier, exciter, generator, and sensor time constants, respectively, while G_a , G_e , G_g , and G_s represent their associated gains. The parameter values listed below, constituting the primary basis of this study, have been selected for AVR system evaluation [19].

$$G_a = 10, G_e = 1, G_g = 1, G_s = 1,$$

$$\tau_a = 0.1 \text{ s}, \tau_e = 0.4 \text{ s}, \tau_g = 1 \text{ s}, \tau_s = 0.01 \text{ s},$$

It is worth noting that the generator gain G_g does not remain fixed but exhibits a load-dependent variability intrinsic to the operating regime of the synchronous machine. Specifically, G_g spans a range from approximately 0.7 under no-load conditions to unity at rated loading [20], a characteristic that directly reflects the nonlinear nature and parametric sensitivity of the generator subsystem. Grounding the analysis in the linearized transfer function representations of the individual AVR components, the aggregate closed-loop transfer function governing the overall AVR system dynamics is expressed as:

$$H(s) = \frac{0.1s + 10}{0.0004s^4 + 0.0454s^3 + 0.555s^2 + 1.51s + 11} \quad (1)$$

For $G_g = 1$, the closed-loop feedback system exhibits the following pole locations:

$$s_{1,2} = -0.5198 \pm 4.6642i, s_3 = -12.4891, s_4 = -99.9711$$

For $G_g = 0.7$, the closed-loop feedback system exhibits the following pole locations:

$$s_{1,2} = -0.7959 \pm 4.017i, s_3 = -11.9284, s_4 = -99.9798$$

A standard tool for assessing closed-loop stability is root locus analysis, which traces the migration of system poles across the complex plane as the feedback gain is varied. Consistent with classical control theory, a system is deemed asymptotically stable if and only if all characteristic poles are confined to the open left-half of the s -plane. The computed pole locations satisfy this condition, thereby validating the asymptotic stability of the AVR closed-loop system under the proposed control architecture.

III. BASICS OF FRACTIONAL CALCULUS

Fractional-order calculus extends the classical integer-order framework of differentiation and integration to the broader domain of arbitrary real — and even complex — orders. The conceptual origins of non-integer derivatives trace back to Leibniz in 1695, with subsequent rigorous mathematical foundations established through the independent contributions of Liouville, Grünwald, Letnikov, and Riemann. The unified differintegral operator ${}_a\mathcal{D}_t^\alpha$, where a and t denote the lower and upper bounds of the operation respectively, and $\alpha \in \mathbb{R}$ governs the order of differentiation or integration, serves as the cornerstone of this generalized calculus. Beyond mere mathematical abstraction, this operator provides a versatile and physically meaningful tool for characterizing the memory effects and hereditary dynamics that pervade a wide class of real-world systems. Its formal definition is given by:

$${}_a\mathcal{D}^\alpha = \begin{cases} d^\alpha/dt^\alpha & \text{for } \Re(\alpha) > 0 \\ \int_a^t (d\tau)^{-\alpha} & \text{for } \Re(\alpha) < 0 \\ 1 & \text{for } \Re(\alpha) = 0 \end{cases} \quad (2)$$

Among the several established definitions of fractional-order derivatives, the Caputo formulation stands as the most prevalent and practically favored in engineering applications, owing to its compatibility with classical integer-order initial conditions. For a sufficiently smooth function $f(t) \in C^{n+1}([0, +\infty), \mathbb{R})$ is given by [2]:

$${}_c\mathcal{D}^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t \frac{f^{(n)}(\xi)}{(t-\xi)^{\alpha-n+1}} d\xi \quad (3)$$

where $\Gamma(x)$ denotes the Gamma function, defined as:

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy \quad (4)$$

Under the assumption of zero initial conditions — a standard premise in linear control analysis — the Laplace transform provides an effective and tractable tool for characterizing fractional-order systems in the frequency domain. Accordingly, the Laplace-domain representation of the Caputo fractional derivative of order $\alpha \in \mathbb{R}^+$ reduces to:

$$\mathcal{L}\{D^\alpha f(t), s\} = s^\alpha F(s) \quad (5)$$

This formulation yields transfer functions parameterized by non-integer powers of the complex variable s , reflecting the infinite-dimensional nature of fractional-order operators.

Since an exact realization of such operators would, in principle, demand an unbounded cascade of poles and zeros, practical deployment — whether in numerical simulation environments or physical hardware implementations — necessitates the use of finite-order rational approximations. These reduced representations are constructed to faithfully replicate the essential frequency-domain behavior of the original fractional operator while maintaining computational tractability. A representative and widely referenced approach is the recursive pole-zero distribution scheme introduced by Charef [21], which achieves accurate rational approximation of fractional powers of s over a prescribed frequency band. This methodology underpins the construction of rational transfer function approximations of fractional operators, expressed in the general form:

$$G(s) = s^\alpha, \quad \alpha \in \mathbb{R}^+ \quad (6)$$

Into a rational function represented as a ratio of polynomials expressed in a factorized form:

$$\hat{G}(s) = \frac{\sum_{i=0}^{n-1} (1 + s/z_i)}{\sum_{i=0}^n (1 + s/p_i)} \quad (7)$$

where the coefficients are determined such that the approximation yields a maximum error of γ dB relative to the baseline frequency-domain magnitude response.

By establishing:

$$a = 10^{\gamma/10(1-\alpha)}, \quad b = 10^{\gamma/10\alpha}$$

The locations of zeros and poles inherent in the resulting rational approximation are determined using the following analytical expressions:

$$z_0 = \omega_c \sqrt{b}, \quad p_0 = az_0, \quad z_i = (ab)^i z_0, \quad p_i = (ab)^i az_0$$

The total number of zeros and poles is directly linked to the specified bandwidth and the admissible approximation error, as defined by the following relation:

$$N = \text{integer} \left\lceil \frac{\log(\omega_{\max}/z_0)}{\log(ab)} + 1 \right\rceil$$

IV. FRACTIONAL-ORDER CONTROLLER

The concept of employing fractional-order controllers for dynamic systems was pioneered by Oustaloup, who introduced the CRONE control strategy [22]. The non integer PID control represents an extended form of the conventional PID architecture, extending its capability through non-integer differentiation and integration orders. This enhanced flexibility allows for finer tuning of closed-loop performance, improved robustness against uncertainties, and superior handling of complex system dynamics. The FOPID controller is mathematically represented by the following transfer function:

$$G_c(s) = \frac{U(s)}{E(s)} = K_p + K_i s^{-\lambda} + K_d s^\mu, \quad \lambda, \mu > 0 \quad (8)$$

The time domain equation for such controller has the form:

$$u(t) = K_p e(t) + K_i \mathcal{D}^{-\lambda} e(t) + K_d \mathcal{D}^\mu e(t) \quad (9)$$

where $u(t)$ denotes the control input and $e(t)$ represents the associated error signal.

Specifically, the FOPID controller encompasses several classical structures as special cases. The parameter pairs $(\lambda, \mu) = (1, 1)$, $(0, 1)$, $(1, 0)$, and $(0, 0)$ corresponds respectively

to the conventional PID, PD, PI, and proportional controllers. As illustrated in Fig.3.

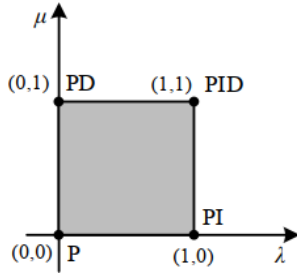


Fig. 3: Visualization of the FOPID controller structure and operation.

The architecture of the FOPID control law is illustrated in Fig. 4.

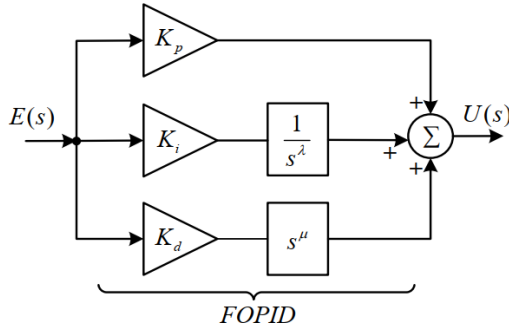


Fig. 4: FOPID controller structure.

V. SINH-COSH ALGORITHM

A. Initialisation case

As with most metaheuristic optimization methods, the SCHO algorithm initializes by generating a population of candidate solutions at random. This population, expressed in (10), is then iteratively updated through the search process, and the best-performing candidate at the end of the iterations is taken as an approximation of the global optimum [16].

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,dim-1} & x_{1,dim} \\ x_{2,1} & \cdots & x_{2,j} & \cdots & x_{2,dim-1} & x_{2,dim} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-1,1} & \cdots & x_{N-1,j} & \cdots & x_{N-1,dim-1} & x_{N-1,dim} \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,dim-1} & x_{N,dim} \end{bmatrix} \quad (10)$$

In this context X denotes the population of candidate solutions generated by (11), where N specifies the population size, $x_{i,j}$ corresponds to the j^{th} element of the i^{th} solution vector, and dim represents the dimensionality of the search space.

$$X = rand(N, dim) \times (ub - lb) + lb \quad (11)$$

where $rand$ represents a random variable uniformly distributed across the interval $[0,1]$, whereas ub and lb correspond to the upper and lower bounds of the decision variables, respectively.

B. Exploration case

The exploration stage is governed by a position-updating mechanism, mathematically expressed through the following relation:

$$X_{(i,j)}^{t+1} = \begin{cases} X_{(i,j)}^{(j)} + r \times W \times X_{(i,j)}^t, & r > 0.5 \\ X_{(i,j)}^{(j)} - r \times W \times X_{(i,j)}^t, & r < 0.5 \end{cases} \quad (12)$$

In this formulation, t represents the current iteration index, while $X_{(i,j)}^{t+1}$ and $X_{(i,j)}^t$ denote the j^{th} element of the i^{th} solution vector at the next and current iterations, respectively.

W_1 denotes the weight coefficient defined as:

$$W_1 = r_3 \times a_1 \times (\cosh r_4 + u \times \sinh r_4 - 1) \quad (13)$$

where r_1 – r_4 are random numbers between 0 and 1. The parameter u is a sensitivity coefficient that regulates the exploration accuracy in the first phase and is fixed at 0.388.

$$a_1 = 3 \times \left(-1.3 \times \frac{t}{Max_iteration} + m \right) \quad (14)$$

Here m is a sensitivity coefficient that governs the exploration accuracy and is set to 0.45 based on experimental results in [16], and t is a coefficient that defines the switching point between the two phases and is fixed at 3.6, as reported in [16].

C. Exploration case

In swarm intelligence (SI) optimization, the exploitation mechanism remains active during the entire evolutionary process. To reinforce search efficiency, this mechanism is structured into two complementary phases, both executed across all iterations. In the first phase, the search is concentrated in the local neighborhood of X , and the corresponding exploitation update is formulated as [16]:

$$X_{(i,j)}^{t+1} = \begin{cases} X_{(i,j)}^{(j)} + r \times W \times X_{(i,j)}^t, & r > 0.5 \\ X_{(i,j)}^{(j)} - r \times W \times X_{(i,j)}^t, & r < 0.5 \end{cases} \quad (15)$$

where W_2 is the weight coefficient given by [16]:

$$W_2 = r_7 \times a_1 \times (\cosh r_8 + u \times \sinh r_8) \quad (16)$$

and r_5 to r_8 are random numbers between 0 and 1. The parameter u retains the same value as in the first exploration phase, being fixed at 0.388.

VI. OPTIMIZATION METHODOLOGY

In the design of controllers, there are several performance criteria. The often-used criteria are the integral of absolute error (IAE) and the integral of squared error (ISE). In this paper, the IAE, expressed in (17), is applied as the objective function due to its strong relevance in voltage regulation problems, particularly for minimizing steady-state tracking error and transient deviations. At each iteration, the voltage error signal given by (18) is utilized to compute the index, while the SCHO algorithm iteratively searches for the optimal set of controller parameters.

$$J = IAE = \int_0^t |e(t)| dt \quad (17)$$

$$e(t) = V_{ref}(t) - V_{out}(t) \quad (18)$$

The parameterization of the FOPID controller consists of five degrees of freedom (5-DoF), specifically three gain terms and two fractional orders, each restricted to specific ranges and investigated within the defined optimization space.

$$K_p^{\min} \leq K_p \leq K_p^{\max}, \quad K_i^{\min} \leq K_i \leq K_i^{\max}, \quad K_d^{\min} \leq K_d \leq K_d^{\max}$$

$$\alpha^{\min} \leq \alpha \leq \alpha^{\max}, \quad \beta^{\min} \leq \beta \leq \beta^{\max}$$

These limits represent the upper and lower bounds (ub and lb) defined in (11). Based on prior literature and practical design considerations, the feasible parameters ranges adopted in this study are:

$$0 \leq K_p \leq 3, \quad 0 \leq K_i \leq 3, \quad 0 \leq K_d \leq 3, \quad 0 \leq \alpha \leq 2, \quad 0 \leq \beta \leq 2$$

It is worth highlighting that incorporating a FOPID controller into the feedback loop may, for certain parameter configurations, jeopardize closed-loop stability. To address this challenge, a constraint-aware optimization strategy is employed in this work. Each candidate solution—representing a unique combination of FOPID parameters—is thoroughly assessed to ensure stability guarantee of the system, with poles located strictly in the left-half s -plane and trajectories converging asymptotically. Any candidate leading to instability is systematically discarded, regardless of its fitness value.

Fig. 5 illustrates the optimization framework employed for tuning the FOPID controller parameters.

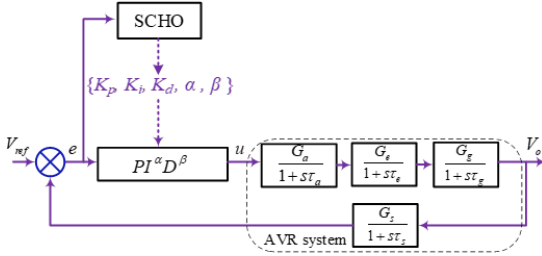


Fig. 5: SCHO-based FOPID optimization framework

VII. SIMULATION RESULTS

This section evaluates the effectiveness of the proposed SCHO-optimized FOPID controller for AVR regulation. The closed-loop performance achieved with the SCHO-based FOPID is systematically compared against conventional controllers, including the PI and standard FOPID schemes and other algorithms-based controllers. As an initial step, and based on the transfer function given in (1), the AVR system is examined in its basic feedback configuration without any external controller. The corresponding response, shown in Fig. 6, reveals oscillatory and weakly damped dynamics, with the feedback loop alone unable to suppress instability. Consequently, the control input u fails to provide adequate stabilization, leading to persistent oscillations and divergence from the reference signal.

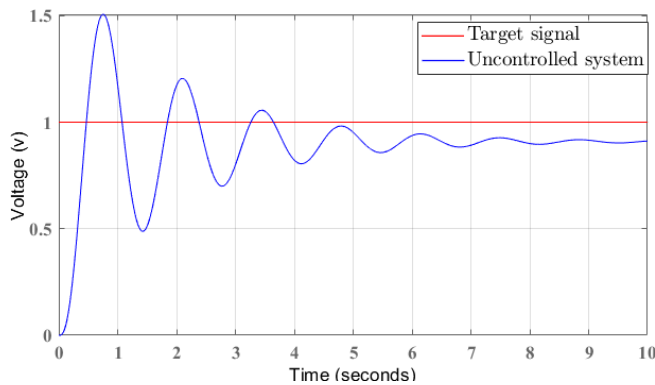


Fig. 6: SCHO-based FOPID optimization framework

Based on the Ziegler–Nichols tuning method, the FOPID controller parameters are initially determined. The fractional orders are approximated using an 8th-order filter ($N = 8$) with a maximum magnitude error of 1 dB ($y = 1$ dB) over the frequency range $\omega \in [10^{-3}, 10^3]$ rad/s. The SCHO algorithm is then employed to optimally refine the FOPID parameters. In SCHO, the population size is set to 30 and the number of iterations to 50, ensuring an effective balance between exploration and exploitation. The optimization process is repeated 20 times to ensure thorough evaluation, and the best

solution among the 20 runs is selected as the controller parameters for the simulations.

A comparative assessment of the AVR terminal voltage response under different control strategies is illustrated in Fig. 7. The performance of the proposed SCHO-optimized FOPID controller is benchmarked against PI and standard FOPID controllers.

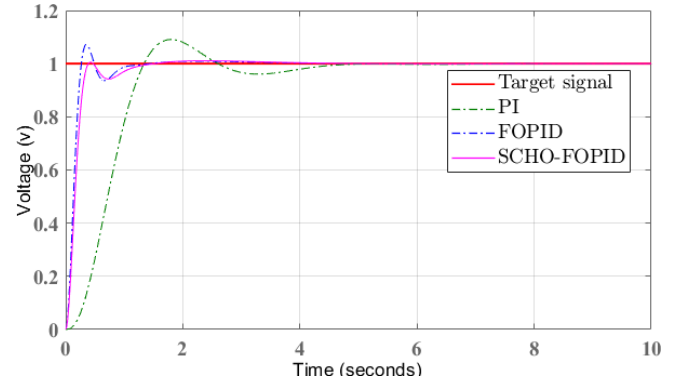


Fig. 7: AVR step response: PI vs. FOPID vs. SCHO-FOPID.

From the closed-loop responses, several critical observations can be drawn. First, the PI controller, while able to regulate the terminal voltage, exhibits a relatively sluggish transient response, marked by a slower rise time, larger overshoot, and extended settling period. These characteristics suggest limited effectiveness in applications requiring rapid stabilization of generator terminal voltage under sudden load disturbances. The standard FOPID controller provides a noticeable improvement by leveraging fractional calculus to enhance the system's memory effect and flexibility in feedback shaping. This results in faster convergence compared to PI control and reduced oscillations. However, the transient performance still suffers from noticeable overshoot and longer settling time, indicating that stability margins remain suboptimal. By contrast, the SCHO-FOPID controller delivers a superior dynamic profile. The closed loop response exhibits: minimal overshoot, rapid rise time and significantly shorter settling time and negligible steady-state error. The SCHO-FOPID reduces overshoot by 86.22% compared with the FOPID controller and by 88.35% compared with the PI controller. Additionally, The SCHO-FOPID reduces settling time by 77.38% compared with the PI controller and by 1.76% compared with the FOPID. The optimal parameter set of the SCHO-FOPID is reported in Table II, while Table III presents a comprehensive comparison of the transient response indices obtained for all controllers.

Table. II

SIMULATION PARAMETERS SETTING USING SCHO

K_p	K_i	K_d	α	β
0.9382	0.6371	0.3742	1.2115	0.9045

Table. III

PERFORMANCE INDICES COMPARISON

Parameters	Controller		
	SCHO-FOPID	FOPID	PI
K_p	0.9382	0.8252	0.1877
K_i	0.6371	1.6051	0.1388
K_d	0.3742	2.1590	-
α	1.2115	0.8541	-
β	0.9045	0.9357	-
Settling time t_s (s)	1.1347	1.1550	5.0087
Rise time t_r (s)	0.3826	0.2436	1.3391
Overshoot (%)	1.2400	9.0000	10.6300

Another comparative analysis is carried out with respect to the state-of-the-art optimization methods, namely ABC and AHA, while uniform constraints were applied across all algorithms: a population size of 50 candidate solution set, a maximum of 100 iterations, and 5 search variables. Each algorithm was run independently 10 times to account for stochastic variability. As shown in Fig. 8, the SCHO-FOPID approach exhibits superior voltage stability within the AVR system, achieving the lowest overshoot and fastest settling time, while ABC-FOPID yields the shortest rise time. Furthermore, SCHO-FOPID provides a smoother transient response with minimal oscillations, demonstrating strong robustness against unmodeled dynamics and load disturbances. These results confirm its improved closed-loop stability and dynamic performance, ensuring higher reliability for AVR applications in renewable energy systems. The key reference-tracking results for all optimization-based FOPID control are listed in Table IV.

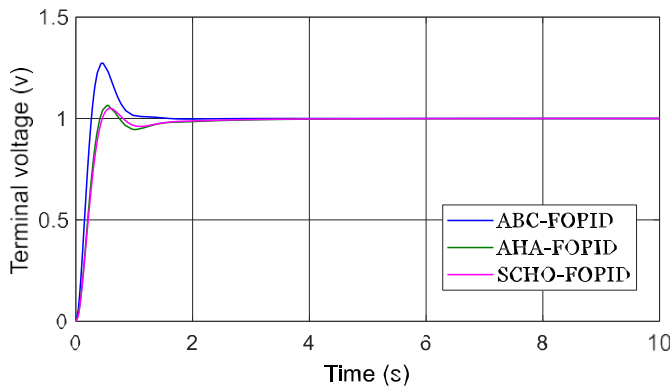


Fig. 8: AVR step response comparison using different FOPID tuning algorithms.

Table IV

Time response specifications	COMPARISON OF DIFFERENT ALGORITHMS		
	Controller		
	SCHO-FOPID	AHA-FOPID	ABC-FOPID
Settling time t_s (sec)	1.4750	1.7680	1.8010
Rise time t_r (sec)	0.4376	0.4096	0.2647
Overshoot (%)	6.0300	6.4400	27.3100

A central requirement of the FOPID controller design is to guarantee closed-loop stability and deliver a well-shaped step response, particularly in the face of abrupt parametric excursions. To this end, a systematic robustness assessment of the SCHO-optimized FOPID-controlled AVR system has been carried out, wherein the plant time constants are individually perturbed across a symmetric range of $\pm 50\%$ around their nominal values in discrete increments of 25%. The resulting closed-loop responses under the proposed control strategy are depicted in Fig. 9. The SCHO-tuned FOPID architecture consistently demonstrates enhanced transient characteristics and sustained resilience against model uncertainty throughout the entire perturbation range. Inspection of the output trajectories reveals that the system response remains confined within a remarkably narrow deviation band, attesting to the exceptional parametric robustness conferred by the proposed controller. These findings collectively confirm the capacity of the SCHO-FOPID scheme to preserve closed-loop stability and regulation quality across the full spectrum of examined parameter variations.

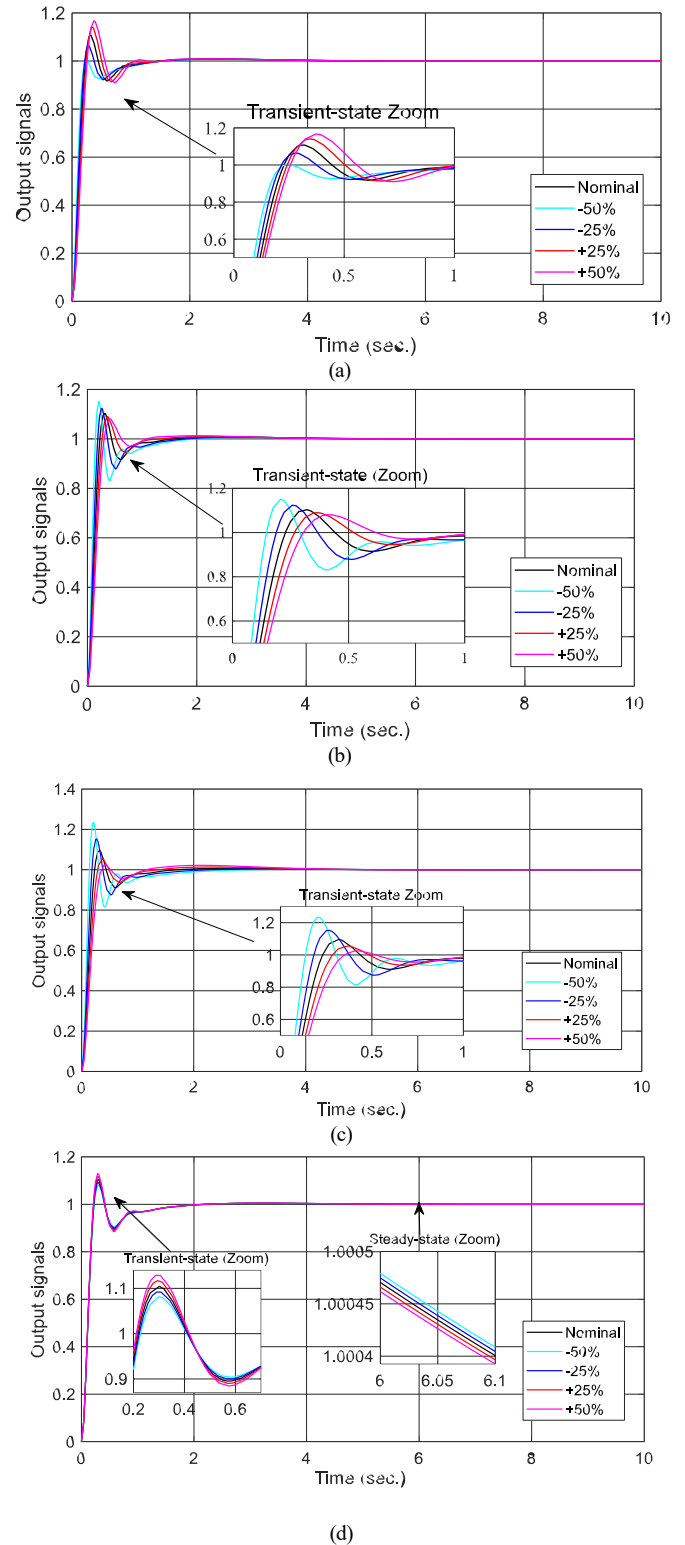


Fig. 9: Robustness analysis of the SCHO-FOPID-controlled AVR under sudden variations in time constant parameters for (a) T_a , (b) T_e , (c) T_g , (d) T_s .

VIII. CONCLUSION

This study presented a systematic and computationally intelligent framework for the optimal synthesis of a Fractional-Order PID (FOPID) controller applied to the Automatic Voltage Regulator (AVR) problem, a critical benchmark for nonlinear power system control under parametric uncertainty. The Sinh-Cosh Optimizer (SCHO), a novel metaheuristic grounded in the complementary analytical dynamics of hyperbolic trigonometric functions, was employed to simultaneously and optimally calibrate the five degrees of freedom (5-DoF) of the FOPID structure, encompassing the proportional, integral, and derivative gains alongside the non-integer integration and differentiation orders.

The synergistic integration of fractional calculus with computational intelligence endows the proposed SCHO–FOPID scheme with a markedly expanded control authority relative to classical integer-order architectures. This expanded parameterization enables the controller to capture memory-dependent and hereditary system dynamics, resulting in superior closed-loop accuracy, enhanced transient shaping, and robust steady-state regulation under diverse operating conditions.

Simulation results unambiguously validated the effectiveness of the proposed strategy in compensating for system nonlinearities, unmodeled dynamics, and abrupt load disturbances. Rigorous comparative benchmarking against classical PID, standard FOPID, and state-of-the-art metaheuristic-tuned controllers (including those based on the Artificial Bee Colony (ABC) and Artificial Hummingbird Algorithm (AHA) optimizers) confirmed the quantitative superiority of the SCHO–FOPID scheme. Specifically, the proposed controller achieved a 19.86% reduction in settling time and a 6.80% decrease in overshoot relative to the AHA-tuned FOPID benchmark, alongside a significant improvement in the ITSE performance index, collectively reflecting accelerated convergence and tighter reference tracking.

Robustness evaluations conducted under structured parametric perturbations — spanning $\pm 50\%$ variations in the AVR plant time constants — further demonstrated the controller's capacity to maintain stable and well-regulated closed-loop behavior across a wide uncertainty envelope, without significant degradation in dynamic performance. These results underscore the resilience of the fractional-order control paradigm when coupled with a well-balanced optimization algorithm.

From a broader perspective, this work reinforces the growing relevance of fractional-order intelligent control in addressing the increasingly complex demands of modern power systems, particularly in renewable energy integration and smart grid contexts, where nonlinearity, intermittency, and parameter variability are inherent operational realities.

Future research directions may encompass the extension of the SCHO–FOPID framework to multi-machine power system configurations, its adaptation for real-time embedded implementation on hardware platforms, and the incorporation of adaptive or self-tuning mechanisms to address time-varying and nonstationary operating environments. The investigation of incommensurate fractional structures and their interaction with advanced stability certification tools also represents a promising avenue for subsequent contributions.

REFERENCES

- [1] W. M. Alghamdi and M. O. Alassafi, "Effective Alternator Voltage Control Based on Computational Intelligence Using Dream Optimizer," 2026.
- [2] E. Bouguenna, S. Ladaci, B. Lekouaghet, W. Merrouche, and M. Benghanem, "Fractional Order PID Controller Design for an AVR System Using the Artificial Hummingbird Optimizer Algorithm," *Int. J. Robust Nonlinear Control*, pp. 1–25, 2025, doi: 10.1002/rnc.7894.
- [3] I. Podlubny, "Fractional-order systems and PI λ D μ -controllers," *IEEE Trans. Autom. Control*. 1999, vol. 44, no. 1, pp. 208–214.
- [4] S. Ladaci, J. J. Loiseau, and A. Charef, "Fractional order adaptive high-gain controllers for a class of linear systems," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 13, no. 4, pp. 707–714, 2008, doi: 10.1016/j.cnsns.2006.06.009.
- [5] E. Bouguenna, S. Ladaci, and W. Merrouche, "Stability and Performance Enhancement of a Step-Down Converter in Photovoltaic Systems Using an Optimization Algorithm-Based Fractional PI λ D δ Controller," pp. 170–178, 2025, doi: 10.1007/978-3-031-80301-7_19.
- [6] B. Hekimoğlu, "Optimal Tuning of Fractional Order PID Controller for DC Motor Speed Control via Chaotic Atom Search Optimization Algorithm," *IEEE Access*, vol. 7, pp. 38100–38114, 2019, doi: 10.1109/ACCESS.2019.2905961.
- [7] S. Ekinici, C. Turkeri, I. Gokalp, Y. Li, and D. Li, "High-performance temperature regulation of nonlinear CSTRs via a hybrid stellar oscillation optimizer and differential evolution-based PID-F control," *Sci. Rep.*, vol. 16, no. 1, 2026, doi: 10.1038/s41598-026-38354-5.
- [8] W. Merrouche, B. Lekouaghet, and E. Bouguenna, "Adolescent Identity Search Algorithm (AISA) for Li-ion Battery Electrical Model Optimization," *Int. J. Artif. Intell. Tools Appl.*, vol. 2, no. 1, 2024.
- [9] A. M. Mosaad, M. A. Attia, N. M. Elbehairy, M. Alruwaili, A. Yousef, and N. M. Hamed, "Enhancing AVR System Stability Using Non-Monopolize Optimization for PID and PIDA Controllers," pp. 1–17, 2025.
- [10] M. S. Ayas and E. Sahin, "FOPID controller with fractional filter for an automatic voltage regulator," *Comput. Electr. Eng.*, vol. 90, no. April 2020, p. 106895, 2021, doi: 10.1016/j.compeleceng.2020.106895.
- [11] Ö. Can, M. Ş. Ayas, and E. Çelik, "Frequency and voltage stability improvement in a two-area thermal power system using a novel controller and RIME optimizer," *Comput. Electr. Eng.*, vol. 125, p. 110434, Jul. 2025, doi: 10.1016/j.compeleceng.2025.110434.
- [12] W. Merrouche, B. Lekouaghet, I. A. E. Boughiout, E. Bouguenna, and M. Benghanem, "Fully Informed Search Algorithm for Estimating the Parameters of Li-Ion Battery Model under UDDS Drive Cycle Profile," *Transp. Res. Procedia*, vol. 84, pp. 275–282, 2025, doi: 10.1016/j.trpro.2025.03.073.
- [13] W. Merrouche, B. Lekouaghet, and E. Bouguenna, "Gaussian Quantum-behaved PSO strategy for Lithium Battery model optimization," in *2023 International Symposium on Quantum Sciences: Applications and Challenges (QSAC2023)*, Algiers, 2023, p. 13.
- [14] T. Joyce and J. M. Herrmann, "A Review of No Free Lunch Theorems, and Their Implications for Metaheuristic Optimisation," 2018, pp. 27–51. doi: 10.1007/978-3-319-67669-2_2.
- [15] E. Bouguenna, S. Ladaci, and B. Taghezouit, "Robust Optimal Fractional-Order PI λ -PD μ Cascade Control Using a PID- Inspired Search Algorithm for DC-DC Converters," *IFAC Pap.*, vol. 59, no. 37, pp. 250–255, 2026, doi: 10.1016/j.ifacol.2026.01.043.
- [16] J. Bai et al., "A Sinh Cosh optimizer," *Knowledge-Based Syst.*, vol. 282, no. September, p. 111081, 2023, doi: 10.1016/j.knosys.2023.111081.
- [17] G. Singh, D. Kumar, P. Aryan, T. R. Chelliah, and G. L. Raja, "Robust and Unified Fractional Order AGC-AVR Scheme for Hydropower Plant," *IEEE Trans. Ind. Appl.*, vol. PP, pp. 1–14, 2026, doi: 10.1109/TIA.2026.3675039.
- [18] E. Bouguenna and S. Ladaci, "Hippopotamus Optimization Algorithm-Based Robust Tilt-FOID Control Design for Linear Time-Invariant Systems in Renewable Energy Applications," *2025 Int. Conf. Artif. Intell. Embed. Syst. Renew. Energy*, pp. 1–6, 2025, doi: 10.1109/AIESRE67541.2025.11470220.
- [19] E. Bouguenna, B. Lekouaghet, and M. Haddad, "Improved Fractional PI α D β Controller for AVR System via a New Optimization Algorithm," *2024 2nd Int. Conf. Electr. Eng. Autom. Control. ICEEAC 2024*, 2024, doi: 10.1109/ICEEAC61226.2024.10576215.
- [20] S. Ekinici and B. Hekimoğlu, "Improved Kidney-Inspired Algorithm Approach for Tuning of PID Controller in AVR System," pp. 39935–39947, 2019.
- [21] A. Charef, H. H. Sun, Y. Y. Tsao, and B. Onaral, "Fractal system as represented by singularity function," *IEEE Trans. Automat. Contr.*, vol. 37, no. 9, pp. 1465–1470, 1992, doi: 10.1109/9.159595.
- [22] A. Oustaloup, F. Levron, B. Mathieu, and F. M. Nanot, "Frequency-band complex noninteger differentiator: Characterization and synthesis," *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.*, vol. 47, no. 1, pp. 25–39, 2000, doi: 10.1109/81.817385.

Thermal Effects of Artificial Lamination Faults in a Three-Phase Transformer Core Using Infrared Thermography

Ehsan Altayef, Sofiane Chiheb, Fateh Anayi, Michael Packianather and Eheda Hassan

Abstract—Transformer core lamination faults can significantly affect transformer performance, reliability, and operational lifetime by increasing core losses and localized heating. This paper investigates the thermal effects of edge burr faults and inter-laminar insulation deterioration. An experimental test rig consisting of a 15 kVA three-phase transformer and a FLIR C2 thermal imaging camera was developed to capture temperature distributions under healthy and faulty operating conditions. Faults were introduced by creating controlled short circuits between laminations with varying affected areas and numbers of shorted laminations. Thermal images were acquired at different magnetic flux densities ranging from 0.5 T to 1.8 T under no-load conditions. The results demonstrate that both fault types produce noticeable temperature rises compared with healthy operation, with the severity of heating increasing as the flux density and fault size increase. Edge burr faults resulted in localized hot spots and temperature rises of up to 10% compared with healthy conditions, while insulation deterioration faults produced temperature increases exceeding 18% for severe fault scenarios at 1.8 T. The findings confirm that infrared thermography provides an effective non-invasive technique for detecting and assessing transformer core lamination faults and can support condition monitoring and early fault diagnosis in power transformers.

Keywords—Power transformer, Transformer core faults, Edge burrs, Lamination insulation degradation, Infrared thermography, Thermal imaging, Condition monitoring.

I. INTRODUCTION

Power transformers are among the most critical and expensive assets in electrical power systems, serving as essential links between generation, transmission, and distribution networks. Their reliable operation is therefore vital for ensuring continuity of supply and maintaining the stability of modern power grids [1]–[6]. Although transformer failures can originate from several sources, defects associated with the magnetic core remain particularly important because they can develop gradually and remain undetected until significant damage has occurred [9], [18].

The transformer core is manufactured from thin grain-oriented electrical steel laminations insulated from one another to suppress eddy currents and minimize core losses [13]. The effectiveness of this design relies heavily on the integrity of the insulation coating between adjacent laminations. During manufacturing, transportation, installation, or long-term operation, the insulation layer may deteriorate or become damaged, leading to unintended electrical contact between laminations.

Similarly, mechanical defects such as edge burrs can create localized short-circuit paths that increase circulating currents within the core [14], [15], [17]. These defects generate excessive heat, increase iron losses, accelerate insulation ageing, and may ultimately compromise transformer efficiency and reliability [18], [19].

Early detection of core lamination faults has attracted considerable attention in recent years. Various diagnostic techniques have been proposed, including no-load current analysis, vibration monitoring, acoustic emission methods, flux measurement, dissolved gas analysis, and frequency-response-based approaches [10]–[12], [25], [26]. While these techniques can provide valuable information regarding transformer condition, they often require complex signal processing, specialised equipment, or indirect interpretation of fault signatures. Furthermore, some methods may have limited sensitivity to localized defects in their early stages [11], [25].

Infrared thermography has emerged as a promising non-contact diagnostic tool for condition monitoring of electrical equipment [7], [8]. Since thermal abnormalities are frequently associated with increased losses and fault development, thermal imaging can provide direct visual evidence of abnormal operating conditions. In transformer cores, localized short circuits between laminations generate hot spots that can be detected through temperature distribution analysis [20]–[24]. Despite the growing interest in thermographic inspection, relatively few experimental studies have systematically investigated the thermal behaviour of different types of artificial lamination faults under controlled operating conditions and varying magnetic flux densities [21]–[24].

This paper presents an experimental investigation into the thermal effects of two common transformer core defects: edge burr faults and inter-laminar insulation deterioration. A dedicated fault-test platform based on a 15 kVA three-phase

Manuscript received June 28, 2026; accepted July 4, 2026.

Ehsan Altayef, Fateh Anayi, and M. Packianather are with School of Engineering Cardiff University, United Kingdom (e.altayef@gmail.com; Anayi@cardiff.ac.uk; PackianatherMS@cardiff.ac.uk.)

Sofiance Chiheb is with National Higher School of Technology and Engineering, Annaba, Algeria (s.chiheb@ensti-annaba.dz)

Eheda Hassan is with University of Hertfordshire, United Kingdom (Eheda222@hotmail.com)

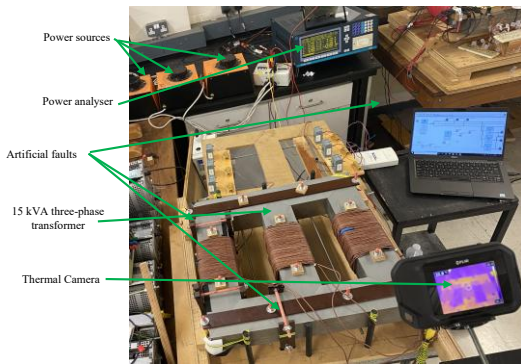
Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.396

transformer was developed to simulate various fault severities by creating controlled short circuits between laminations. Thermal images were acquired using a FLIR infrared camera under healthy and faulty conditions for flux densities ranging from 0.5 T to 1.8 T. The objective is to quantify the relationship between fault severity, magnetic flux density, and temperature rise within the transformer core, thereby evaluating the effectiveness of thermal imaging as a diagnostic tool for detecting and assessing lamination-related faults.

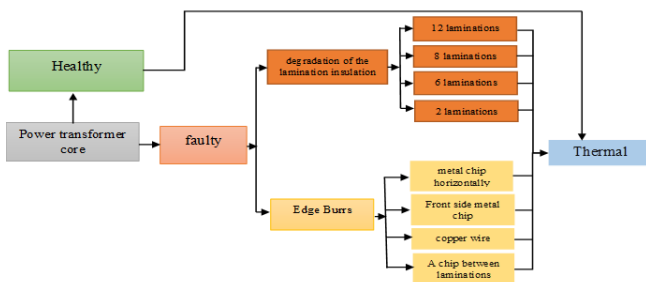
The contributions of this work are threefold. First, a comprehensive experimental dataset of thermal responses for multiple artificial lamination fault scenarios is established. Second, the influence of fault size and flux density on transformer core temperature is analysed quantitatively. Third, the study demonstrates the capability of infrared thermography to distinguish between healthy and faulty operating conditions, providing valuable insight for future transformer condition monitoring and fault diagnosis systems. The experimental framework and fault scenarios investigated in this study build upon previous investigations of transformer core lamination faults and their effects on transformer performance [27]–[29]. Recent advances in transformer fault diagnosis and electrical machine analysis have also provided additional motivation for the present work [30]–[33].

II. EXPERIMENTAL SET-UP AND SAMPLE

As complimentary studies for previous fault and estimation works (e.g., [30]–[33]), the paper treat transformer core faults. This section describes the data acquisition system and experimental procedure employed to investigate the effects of lamination faults in power transformers. The experimental fault-testing platform was developed by the Magnetics Group at the Cardiff School of Engineering to facilitate controlled fault simulation and thermal analysis. The system enables the identification of hotspot regions within the transformer core through thermal imaging. Figure 1 presents a photograph of the experimental setup together with a schematic representation of the measurement system..



(a) Experimental setup



(b) Faulty operation diagram.
Fig. 1: Experimental Set-up

The experimental test rig comprises several key components, including a 15 kVA three-phase power transformer, specially designed clamping fixtures to secure the transformer during the introduction of edge burr faults, and a rotary tool used to remove approximately 400 mm² of insulation from the transformer core laminations to simulate the second fault condition investigated in this study. The magnetic flux density was determined from the measured voltage and current signals acquired using a power analyser connected to the transformer. A thermal imaging camera was employed throughout the experiments to capture the temperature distribution of both healthy and faulty transformer cores, enabling a comparative thermal analysis.

The transformer core is constructed from a stack of 520 grain-oriented electrical steel laminations. It features a dual-window core structure with window dimensions of 320 mm × 120 mm and overall core dimensions of 540 mm × 520 mm. Furthermore, both the primary and secondary windings are uniformly wound around the core limbs, each consisting of 50 turns of insulated copper conductor with a cross-sectional area of 1.5 mm².

III. PROPOSED FAULTS DETAILS

A. Artificial Lamination Faults

In this study, two common transformer core faults were investigated: edge burrs and interlaminar insulation deterioration. These faults are among the most frequently occurring defects in transformer cores and can significantly affect transformer performance. To evaluate their impact, no-load tests were conducted, as no-load current measurements provide an indication of the transformer's no-load (iron) losses. During these tests, the transformer primary winding was energised with its rated voltage of 220 V, while the secondary winding remained open-circuited.

To assess the thermal effects of the investigated faults, the transformer was first examined under healthy operating conditions to establish a baseline for comparison. During this stage, the three-phase voltage and current were recorded at magnetic flux densities of 0.5, 0.8, 1.0, 1.5, 1.7, and 1.8 T. Each experiment was repeated three times to ensure the accuracy, repeatability, and reliability of the measured data. Furthermore, the data for each fault condition were collected on separate days, allowing sufficient time for the transformer core to return to ambient temperature before subsequent measurements. This procedure minimised the influence of residual heating on the experimental results. The following sections describe each fault condition in detail.

B. Edge Burr Fault

To simulate the edge burr fault, an artificial short circuit was introduced between adjacent laminations of the transformer core. Four fault scenarios were considered, each representing a different number of short-circuited laminations (i.e., different affected areas). Prior to fault implementation, baseline measurements were obtained under healthy operating conditions over a range of magnetic flux densities from 0.5 T to 1.8 T to establish a reference for comparison.

Following the baseline measurements, edge burr faults were introduced at various locations on the transformer core according to the predefined fault scenarios. A custom-designed clamping device, shown in Figure 2, was developed to fit the experimental transformer core used in this study. The clamping mechanism ensured firm and consistent electrical contact

between the artificial burr material and the edges of the laminated core stack, thereby providing repeatable and reliable fault conditions throughout the experiments.

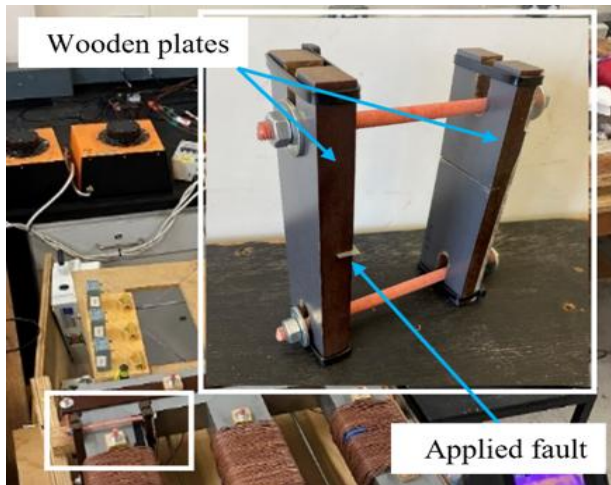


Fig. 2: Clamping device used for the laminations fault fixation

As illustrated in Figure 2, the clamping device consists of two wooden plates fastened using two plastic bolts, which apply uniform pressure to the copper foil placed on both sides of the laminated core stack. Since both wood and plastic are non-conductive and exhibit negligible electromagnetic interference, their influence on the experimental measurements can be considered insignificant. This assumption is consistent with previous studies, which have similarly neglected the effect of the clamping device on the experimental results (e.g., [19], [27]).

Four edge burr fault scenarios were implemented by varying the number of short-circuited laminations and the corresponding contact area, as illustrated in Figure 3. In the first scenario (Figure 3(a)), a thin metal strip was inserted to create a short circuit between two adjacent laminations within the 520-lamination transformer core. The contact area measured 45 mm in length and 0.5 mm in width, corresponding to an affected area of 22.5 mm².

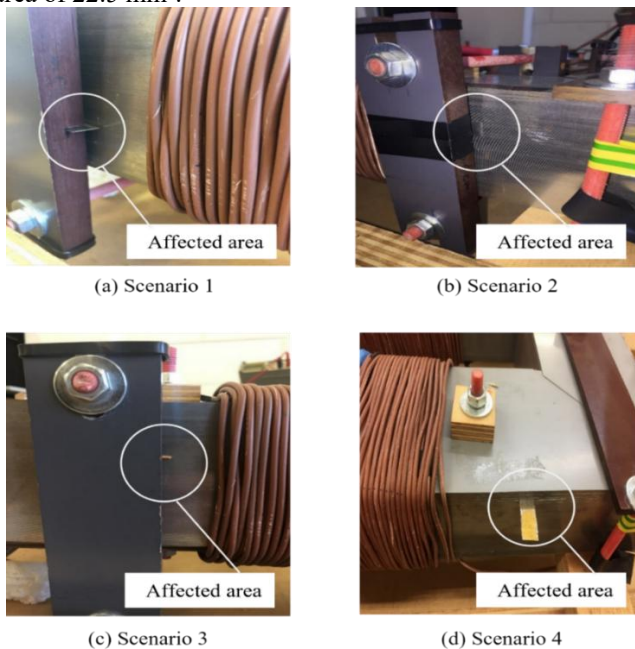


Fig. 3: Artificial edge burr lamination faults (a) scenario 1, (b) scenario 2, (c) scenario 3 and (d) scenario 4

the second scenario (Figure 3(b)), a wider metal strip measuring 45 mm × 10 mm was employed, producing a contact area of 450 mm². This configuration increased the number of short-circuited laminations from two to approximately 33, thereby

representing a more severe edge burr fault.

For the third scenario (Figure 3(c)), a copper wire with a diameter of 1.3 mm was used to establish a short circuit across four adjacent laminations. The resulting contact area measured approximately 45 mm × 1.3 mm (58.5 mm²).

The fourth fault scenario, shown in Figure 3(d), consisted of a localized short circuit between two laminations with a contact area of 9 mm × 15 mm (135 mm²). This configuration was included to investigate the influence of contact geometry on the thermal behaviour of the transformer core.

C. Lamination Insulation Fault

Transformer cores are constructed from thin electrical steel laminations to minimise eddy current losses and improve operating efficiency. Each lamination is coated on both sides with a thin inorganic insulating layer, typically 1–3 μm thick, to prevent direct electrical contact between adjacent laminations. However, deterioration of the inter-laminar insulation may occur due to several factors, including ageing of the insulating coating, mechanical damage caused by external objects, and overheating resulting from winding failures or other abnormal operating conditions [29].

To investigate the thermal effects of inter-laminar insulation deterioration, artificial insulation faults were introduced into the transformer core, and thermal images were captured to assess the resulting temperature distribution. The fault was created by carefully removing the insulating coating from the selected laminations and inserting a copper strip between them to maintain electrical contact, thereby simulating insulation breakdown. The fault preparation procedure is illustrated in Figure 4.

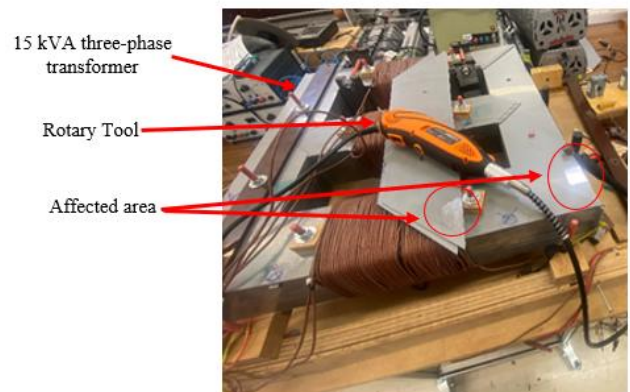


Fig. 4: Experimental setup used for the laminations fault analysis

To simulate inter-laminar insulation deterioration, artificial insulation damage was introduced on the opposing surfaces of selected transformer core laminations. Fault conditions involving short circuits between 2, 6, 8, and 12 laminations were investigated at magnetic flux densities of 0.5, 1.0, 1.5, 1.7, and 1.8 T. The insulation damage was created by removing approximately 40 mm² of the insulating coating using a rotary tool, thereby establishing electrical contact between adjacent laminations. To evaluate the influence of fault location on the transformer's thermal behaviour, the artificial faults were introduced at different positions within the transformer core.

IV. HEALTHY OPERATION MODE

This section presents the experimental results obtained from the thermal analysis of the transformer under healthy and faulty operating conditions. Thermal images were captured at various

magnetic flux density levels to evaluate the influence of the investigated lamination faults on the transformer core temperature distribution. A FLIR C2 thermal imaging camera was used to record all thermal images. For each fault scenario and flux density level, six thermal images with a resolution of 80×60 pixels were captured at 15-minute intervals throughout the experimental period. This acquisition procedure ensured consistent monitoring of the temperature evolution under each operating condition. The specifications of the FLIR C2 thermal camera are summarised in Table 1.



Fig. 5: Thermal camera

The specifications of the camera are presented in table 1.

TABLE 1 SPECIFICATIONS OF THE CAMERA

PARAMETER	VALUE
1 Focal length	1.54 mm
2 Size (L*W*H)	124.46*78.74*12.44 mm
3 IR sensor	80*60 (4,800 measurement pixel)
4 Operating temperature range	-10°C to +50°C
5 Storage temperature range	-40°C to +70°C
6 Digital camera	640*480 pixel
7 Image frequency	9 Hz
8 Accuracy	$\pm 2^\circ\text{C}$
9 Thermal sensitivity	$< 0.10^\circ\text{C}$

The thermal analysis was performed by first capturing thermal images of the transformer core under healthy operating conditions to establish a reference for comparison with the faulty conditions. These reference images provided a baseline for evaluating the temperature variations caused by the introduced lamination faults.

Although the FLIR C2 thermal camera has a relatively low image resolution compared with more recent thermal imaging systems, it was sufficient to identify the temperature distribution and hotspot locations associated with the investigated faults. Since the induced faults were located at the edges of the transformer core laminations, thermal images were captured from the side profiles of the transformer rather than the front surface. This viewing angle provided a clearer representation of the fault-affected regions and enabled more accurate observation of the corresponding thermal patterns, as illustrated in Figure 6.

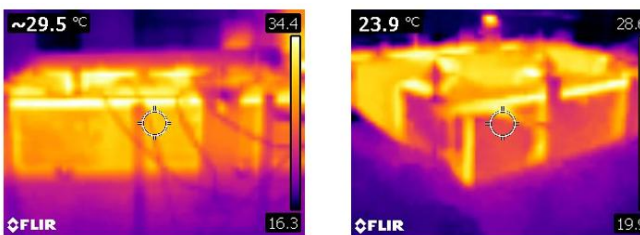


Fig. 6: The caught image on the profiles of the transformer

For improved visualisation, Figure 7 presents the thermal images of the transformer core under healthy operating

conditions prior to the introduction of any faults. These images serve as a reference for evaluating the thermal effects of the investigated lamination faults. The maximum and minimum temperatures obtained from each thermal image provide a clear representation of the transformer core temperature distribution at different magnetic flux density levels. These temperature values were subsequently extracted as input features for the fault detection and classification process described in the following section.

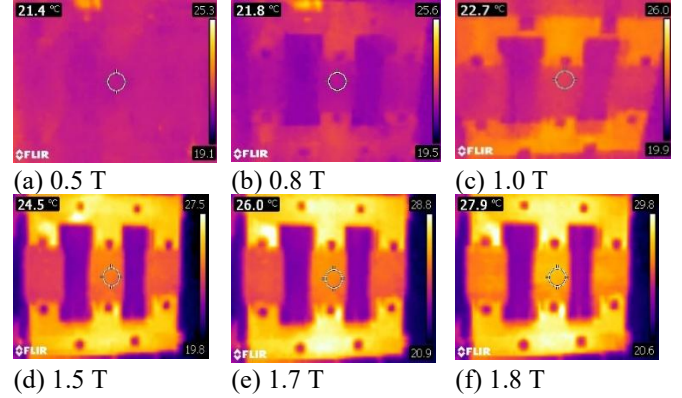


Fig. 7: Healthy operation mode results

For the healthy operating condition of the transformer core, it is observed that magnetic flux density has a significant influence on the temperature rise in the no-load transformer. At low flux density levels, as shown in Figure 7(a), the core temperature remains relatively low, with maximum and minimum values of approximately 25.3°C and 19.1°C , respectively. In contrast, at the highest flux density of 1.8 T (Figure 7(f)), a noticeable increase in temperature is observed, with maximum and minimum values rising to approximately 29.8°C and 23.3°C , respectively. This trend clearly indicates that higher magnetic flux densities lead to increased core heating under no-load conditions.

V. FAULTY OPERATION MODE

A. Edge Burrs fault

This section presents the thermal imaging results for the first type of fault considered in this study, namely the edge burr fault. This fault was implemented at different scenarios and locations within the transformer core to evaluate its thermal impact. Figures 8 and 9 illustrate representative thermal images obtained at magnetic flux densities of 0.5 T and 1.8 T for fault scenarios 1 and 4, respectively. These figures are provided to demonstrate the influence of magnetic flux density on the temperature distribution of the transformer core under edge burr fault conditions.

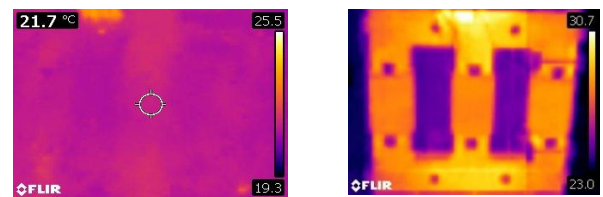


Fig. 8: Scenario 1 faulty operation mode results

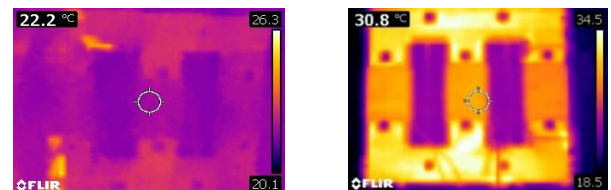


Fig. 9: Scenario 4 faulty operation mode results

The obtained thermal images indicate a clear difference between the temperature distribution under edge burr fault

conditions and that observed during healthy operation. This confirms that edge burr faults have a significant impact on the transformer core temperature and, consequently, on overall transformer performance. In general, an increase in temperature is observed across all fault scenarios.

For example, at a magnetic flux density of 1.8 T, the temperature under healthy conditions is approximately 1 °C lower than that observed in fault scenario 1. A more pronounced increase is observed in scenario 4, where the temperature is approximately 5 °C higher than the healthy condition. Furthermore, scenario 4 exhibits an increase of around 4 °C compared with scenario 1 at the same flux density. These results demonstrate that the rate of temperature rise becomes more significant with increasing magnetic flux density and a higher number of affected laminations within the transformer core.

Edge burr faults lead to increased core temperature, which in turn contributes to higher core losses and increased current losses. This temperature rise may also degrade the properties of the insulating materials, thereby reducing transformer efficiency and operational reliability. For this reason, further investigation of such faults in high-capacity transformers is strongly recommended.

Overall, the thermal imaging results show that edge burr faults can increase transformer core temperature by approximately 3.2% in scenario 1 compared with healthy conditions, and up to approximately 16.1% in scenario 4. In addition, scenario 4 exhibits a temperature increase of about 12.5% relative to scenario 1, indicating a clear relationship between fault severity and thermal response. These findings confirm that both magnetic flux density and the extent of lamination short-circuiting play a critical role in determining the thermal behaviour of the transformer core.

Lamination Insulation Faults

In this section, the thermal imaging results for the second type of fault, namely inter-laminar insulation degradation, are presented. This fault was introduced by selectively removing the insulating coating from the transformer core laminations, thereby creating electrical contact between adjacent layers. Figure 10 illustrates the corresponding thermal images of the transformer core under fault conditions.

The results correspond to short-circuit conditions involving 2, 6, 8, and 12 affected laminations, evaluated at magnetic flux densities of 0.5 T and 1.8 T. It should be noted that the influence of fault location on the position of the inter-laminar short circuit has been previously investigated using current signal analysis in [27]. In this study, the same fault placement strategy and positioning methodology were adopted to ensure consistency and comparability of results.

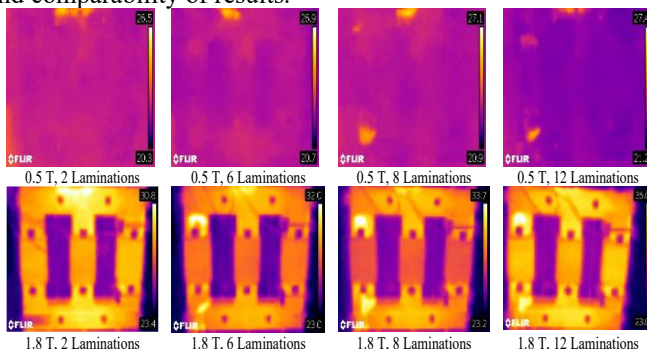


Fig. 10: Faulty operation mode results at 0.5 and 1.8 T

From the results presented in this figure, it is evident that the size of the fault, defined by the number of short-circuited laminations, has a significant influence on the transformer core temperature. The thermal measurements indicate that the temperature increases as the number of affected laminations increases, as illustrated in the figure. Specifically, the figure presents the maximum and minimum temperature values as a function of the number of affected laminations at magnetic flux densities of 0.5 T and 1.8 T.

The results demonstrate that the temperature rise associated with inter-laminar insulation faults is strongly dependent on both the number of shorted laminations and the applied flux density. For fault cases involving 2, 6, 8, and 12 short-circuited laminations, the highest temperatures are observed at a flux density of 1.8 T, with increases exceeding 9 °C compared with the corresponding values at 0.5 T for the 12-lamination fault scenario. For the case of two shorted laminations, the temperature increases by more than 4 °C across the flux density range from 0.5 T to 1.8 T. In addition, temperature rises of approximately 6 °C are observed for the six- and eight-lamination fault cases over the same flux density range.

Overall, the transformer core temperature under insulation degradation faults increases with both the severity of the fault (number of short-circuited laminations) and the magnetic flux density. This pronounced temperature rise may lead to increased core losses, particularly eddy current losses, which can reduce transformer efficiency and, in severe cases, contribute to operational failure.

Finally, the thermal data acquired using the FLIR C2 camera were stored in the device's internal memory and subsequently transferred to a personal computer for post-processing and feature extraction.

VI. CONCLUSION

This paper has presented an experimental investigation into the thermal behaviour of a three-phase transformer core under healthy and faulty lamination conditions using infrared thermography. Two common transformer core defects, namely edge burr faults and lamination insulation deterioration, were artificially introduced into a 15 kVA transformer core and analysed at different magnetic flux densities ranging from 0.5 T to 1.8 T.

The thermal imaging results demonstrated that both fault types cause significant temperature increases compared with healthy operating conditions. For edge burr faults, the transformer core temperature increased with both the flux density and the extent of the short-circuited lamination area. The largest edge burr scenario produced temperature rises of approximately 5 °C above healthy conditions at 1.8 T. Similarly, insulation deterioration faults generated progressively higher temperatures as the number of short-circuited laminations increased. The most severe insulation fault scenario, involving twelve affected laminations, produced temperature increases exceeding 9 °C at high flux density.

The experimental findings indicate that lamination faults create additional eddy-current paths, leading to increased core losses and localized hot spots that may accelerate insulation ageing and reduce transformer efficiency and reliability. Furthermore, the results confirm that thermal imaging is a practical, non-invasive, and effective diagnostic technique for identifying transformer core lamination defects and evaluating their severity.

Future work will focus on advanced thermal image processing and machine learning techniques for automatic fault detection, classification, and severity assessment, as well as extending the investigation to larger power transformers operating under practical loading conditions.

REFERENCES

- [1] P. Chen, Y. Huang, F. J. Zeng, Y. Jin, X. Zhao, and J. Wang, "Review on insulation and reliability of dry-type transformer," in *Proc. IEEE Sustainable Power and Energy Conference (iSPEC)*, 2019, pp. 398–402.
- [2] X. Duan, T. Zhao, J. Liu, L. Zhang, and L. Zou, "Analysis of winding vibration characteristics of power transformers based on the finite-element method," *Energies*, vol. 11, no. 9, 2018.
- [3] C. Jettanasen and A. Ngaopitakkul, "A novel probabilistic neural network-based algorithm for classifying internal fault in transformer windings," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 8, no. 2, pp. 123–131, 2013.
- [4] F. Yuan, J. Guo, Z. Xiao, B. Zeng, W. Zhu, and S. Huang, "A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine," *Energies*, vol. 12, no. 5, 2019.
- [5] B. E. Lee, J. W. Park, P. A. Crossley, and Y. C. Kang, "Induced voltages ratio-based algorithm for fault detection, and faulted phase and winding identification of a three-winding power transformer," *Energies*, vol. 7, no. 9, pp. 6031–6049, 2014.
- [6] M. Gutten, D. Korenciak, M. Kucera, R. Janura, A. Glowacz, and E. Kantoch, "Frequency and time fault diagnosis methods of power transformers," *Measurement Science Review*, vol. 18, no. 4, pp. 162–167, 2018.
- [7] A. Sedighi, A. Kafiri, M. Shahnazari, M. R. Sehati, and F. Behdad, "Aging assessment of distribution transformers based on thermal imaging," in *Proc. IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 2019.
- [8] A. S. N. Huda and S. Taib, "Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment," *Applied Thermal Engineering*, vol. 61, no. 2, pp. 220–227, 2013.
- [9] R. Murugan and R. Ramasamy, "Failure analysis of power transformer for effective maintenance planning in electric utilities," *Engineering Failure Analysis*, vol. 55, pp. 182–192, 2015.
- [10] D. A. K. Pham, T. M. T. Pham, H. Borsi, and E. Gockenbach, "A new method for purposes of failure diagnostics and FRA interpretation applicable to power transformers," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 6, pp. 2026–2034, 2013.
- [11] R. Khalili Senobari, J. Sadeh, and H. Borsi, "Frequency response analysis (FRA) of transformers as a tool for fault detection and location: A review," *Electric Power Systems Research*, vol. 155, pp. 172–183, 2018.
- [12] M. Bagheri, A. Zolavarani, and S. Nezhivenko, "Transformer fault condition prognosis using vibration signals over cloud environment," *IEEE Access*, vol. 6, pp. 9862–9874, 2018.
- [13] S. B. Lee, G. B. Kliman, M. R. Shah, W. T. Mall, N. K. Nair, and R. M. Lusted, "An advanced technique for detecting inter-laminar stator core faults in large electric machines," *IEEE Transactions on Industry Applications*, vol. 41, no. 5, pp. 1185–1193, 2005.
- [14] A. Kedous-Lebouc, B. Cornut, J. C. Perrier, P. Manfè, and T. Chevalier, "Punching influence on magnetic properties of the stator teeth of an induction motor," *Journal of Magnetism and Magnetic Materials*, vols. 254–255, pp. 124–126, 2003.
- [15] P. Handgruber, A. Stermecki, O. Biro, and G. Ofner, "Evaluation of interlaminar eddy currents in induction machines," in *Proc. IECON Annual Conference of the IEEE Industrial Electronics Society*, 2013, pp. 2792–2797.
- [16] M. Bigdeli, M. Vakilian, and E. Rahimpour, "Transformer winding faults classification based on transfer function analysis by support vector machine," *IET Electric Power Applications*, vol. 6, no. 5, pp. 268–276, 2012.
- [17] M. B. Aimoniotis and A. J. Moses, "Evaluation of induced eddy currents in transformer sheets due to edge-burrs, employing computer aided design programs," in *Proc. Athens Power Tech International Power Conference*, vol. 2, pp. 847–850, 1993.
- [18] E. Altayef, F. Anayi, and M. Packianather, "Power transformer fault detection methods and effects of edge burrs fault: A review," in *Proc. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022.
- [19] R. Mazurek, H. Hamzehbahmani, A. J. Moses, P. I. Anderson, F. J. Anayi, and T. Belgrand, "Effect of artificial burrs on local power loss in a three-phase transformer core," *IEEE Transactions on Magnetics*, vol. 48, no. 4, pp. 1653–1656, 2012.
- [20] V. M. Shiljkut, "Remote monitoring of power transformers thermal image," pp. 12–15, 2012.
- [21] Y. Sun et al., "A temperature-based fault pre-warning method for the dry-type transformer in the offshore oil platform," *International Journal of Electrical Power & Energy Systems*, vol. 123, 2020.
- [22] E. Gockenbach, P. Werle, and H. Borsi, "Monitoring and diagnostic systems for dry type transformers," in *Proc. IEEE International Conference on Conduction and Breakdown in Solid Dielectrics*, 2001, pp. 291–294.
- [23] H. Li, "Thermal fault detection and diagnosis of electrical equipment based on the infrared image segmentation algorithm," *Advances in Multimedia*, vol. 2021, pp. 1–7, 2021.
- [24] A. K. Al-Musawi, F. Anayi, and M. Packianather, "Three-phase induction motor fault detection based on thermal image segmentation," *Infrared Physics & Technology*, vol. 104, Art. no. 103140, 2020.
- [25] M. Bagheri, M. Naderi, and T. Blackburn, "Advanced transformer winding deformation diagnosis: Moving from off-line to on-line," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 6, pp. 1860–1870, 2012.
- [26] J. Liu, Z. Zhao, C. Tang, C. Yao, C. Li, and S. Islam, "Classifying transformer winding deformation fault types and degrees using FRA based on support vector machine," *IEEE Access*, vol. 7, pp. 112494–112504, 2019.
- [27] E. Altayef, F. Anayi, M. S. Packianather, and O. Kherif, "On the effects of lamination artificial faults in a 15 kVA three-phase transformer core," *IEEE Access*, vol. 10, pp. 19348–19355, 2022.
- [28] E. Altayef, F. Anayi, M. Packianather, Y. Benmahamed, and O. Kherif, "Detection and classification of lamination faults in a 15 kVA three-phase transformer core using SVM, KNN and DT algorithms," *IEEE Access*, vol. 10, pp. 50925–50932, 2022.
- [29] E. Altayef, F. Anayi, and M. Packianather, "Experimental investigation on impacts of insulation damage fault between laminations of power transformers," in *Proc. 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022.
- [30] O. Kherif et al., "A New Multi-Path Hybrid Classifier for Transformer Oil Fault Diagnostic," *ENP Engineering Science Journal*, vol. 5, no. 2, Dec. 2025.
- [31] A. Khelafi, A. Djebli, O. Touhami, and R. Ibtouen, "Analysis of Multiphase Electrical Systems According to the Number of Phases," *ENP Engineering Science Journal*, vol. 4, no. 2, Dec. 2024.
- [32] F. Leguebedj and D. Boukhetala, "Estimation of Synchronous Machine Parameters by Stand Still Frequency Responses Testing," *ENP Engineering Science Journal*, vol. 3, no. 2, Dec. 2023.
- [33] O. Kherif, T. Zebbadji, Y. Gherbi, M. L. Azzouze, and M. Teguar, "Simplified Diagnosis Method for CHBMs under Open-Circuit Switch or Battery Failure," *ENP Engineering Science Journal*, vol. 1, no. 2, Dec. 2021.

Review of ISO 15118-20 with TLS1.3 for Secure and Trustworthy EV Charging Infrastructure

Akram Salem, Mian Hammad Nazir, Leshan Uggalla, Abdulla Rahil and Khayri Abu Sayf

Abstract—Electric Vehicle Supply Equipment (EVSE) and the communication protocols that support it are essential to modern electric mobility, yet they remain vulnerable to significant cyber-physical risks. These risks largely arise from fragmented security adoption and the inconsistent use of Transport Layer Security (TLS). This paper reviews and analyses the security enhancements introduced by ISO 15118-20, with particular emphasis on the requirement for mandatory, mutually authenticated TLS 1.3 to secure end-to-end electric vehicle charging communication. Through an examination of protocol message flows and a synthesis of documented vulnerabilities, the study shows how stronger cryptographic enforcement, certificate-based authentication, and encrypted control signalling reduce the risks of replay, impersonation, man-in-the-middle, and backend exploitation attacks. The analysis also highlights the cyber-physical consequences of communication failures, demonstrating how protocol-level disruptions can result in wider operational risks. As a promoting solution, the paper concludes that the consistent implementation of ISO 15118-20, supported by effective certificate lifecycle management, is essential for developing resilient and trustworthy EV charging infrastructure. While additional advanced security approaches may further strengthen system robustness, future research is required to address additional technical, operational, and environmental factors not covered in this study.

Keywords— EV Charging, Cybersecurity, ISO 15118, TLS, and Cyber-attack.

NOMENCLATURE

CSMS	Charging Station Management System
EV	Electric Vehicle
EVCS	Electric Vehicle Charging System
EVSE	Electric Vehicle Supply Equipment
HSM	Hardware Security Modules
OCPP	Open Charge Point Protocol
PnC	Plug and Charge
PKI	Public Key Infrastructure
SCADA	Supervisory Control and Data Acquisition
SoC	State of Charge
TLS	Transport Layer Security
V2G	Vehicle-to-Grid

I. INTRODUCTION

EVSE and EVs are fundamental components of sustainable transportation systems. Their widespread deployment introduces significant security challenges that must be addressed to ensure safe operation, reliable grid integration, and protection of sensitive data [1]. These challenges encompass personal and corporate privacy, secure financial transactions, and the stability of the electric grid. This may make cybersecurity a critical requirement for modern EV charging infrastructures [2].

Recent studies (e.g., [2]–[4]) reveal alarming security gaps in existing EV charging infrastructure. In [2], for instance, it was found that approximately 84% of surveyed charging stations do not implement TLS. This resulted, among others, in preventing compatibility with newer versions of the ISO 15118 standard

and exposing systems to severe cyber vulnerabilities. For instance, the ISO 15118 protocol, which defines communication between EVs and EVSE, has evolved considerably to address these issues. Authors in [5] showed that earlier versions, such as ISO 15118-2, treated TLS as optional, limiting their effectiveness against advanced cyber threats. Different versions of the protocol ISO 15118-2 have been flown where some of them have become mandatory with TLS and more authentication during charging session [6], [7].

The latest standard, ISO 15118-20, represents a major advancement by mandating mutually authenticated TLS, significantly enhancing the security posture of EV charging communications. TLS ensures authentication, confidentiality, and data integrity through PKI, as demonstrated in prior research (e.g., [1], [5], [8]). This evolution enables secure services such as PnC where all V2G communications occur within a TLS-encrypted channel [6]. Such improvement thereby helps with reducing vulnerabilities and improving overall system reliability.

Despite these progresses, the rapid growth of EV adoption increases exposure to advanced cyber-attacks, highlighting the need for stronger cybersecurity mechanisms. In response to this challenge, upgrading protocols has been considered (e.g., [5], [8], [9]). In [9], for instance, some versions of OCPP have demonstrated improvements in authentication, confidentiality, and integrity. Additionally, the use of TLS 1.3 in conjunction with ISO 15118-20 has been studied in [1], [9]. However, increasing performance demands, reduced latency requirements, and heightened resilience against emerging threats necessitate further enhancements to existing security frameworks as explained in [10].

To overcome the listed issues, this paper focuses on the security enhancements introduced in ISO 15118-20, with particular emphasis on mandatory mutual TLS authentication and strengthened encryption mechanisms. A review of existing vulnerabilities in EV charging infrastructure has been conducted. The results have been structured to examine the

Manuscript received December 25, 2025; accepted May 5, 2026.

A. Salem, M.H. Nazir, L. Uggalla, A. Rahil and K. Abu Sayf are with the Department of Electric Engineering and Computer Science, University of South Wales, Cardiff, United Kingdom (e-mail: akram.salem@southwales.ac.uk).

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.358

role of TLS and PKI in securing V2G communications as well as evaluate and investigate the impact of the latest standard on addressing current and future cybersecurity challenges. Discussion and analyses have been included to summarise vulnerabilities and associated impacts with the aim of explaining the effectiveness of protocols in safeguarding EVSE and focusing secure, scalable, and reliable electric mobility solutions.

After an introductory part on cybersecurity research in EVSE charging systems, the remainder of this paper is organized into four main sections that together provide a comprehensive review of the literature. The first section presents a brief overview of vulnerabilities in EVSE. The second and third sections examine, respectively, the impacts of cyberattacks on EVSE cybersecurity and emerging cyber threats targeting EVSE infrastructure. The fourth section focuses on the classification of EVSE cyberattacks and corresponding defence mechanisms. An additional section is devoted to the discussion and critical analysis of related work. Finally, the paper concludes with a summary of findings and outlines directions for future research.

II. CYBERSECURITY IN EVSE CHARGING SYSTEMS

The EVSE communication structure can be grouped into three main components: EV, EVSE, and CSMS. To assess and understand cybersecurity weaknesses and vulnerabilities in EVSE systems, the interactions between EV, EVSE, and CSMS must be considered. Figure 1 presents a high-level representation of the EVSE communication architecture, distinguishing between frontend communication governed by ISO 15118 and other communication paths.

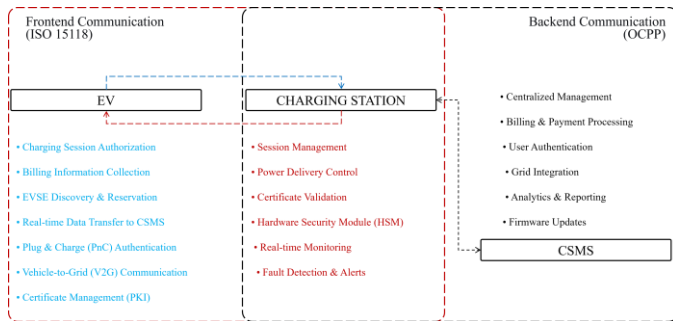


Fig. 1 Representation of EVSE Communication Structure

As shown in the figure, communication between the EV and EVSE is bidirectional. One direction includes authentication, billing, and related procedures, while the other covers session management, power delivery control, and other operational functions. The EVSE communicates with the CSMS to enable centralized management, payment processing, and additional functionalities such as grid integration and firmware updates. Security risks may occur at any communication layer and can affect one or multiple functionalities. The severity of the impact depends on factors such as the type of attack and the success of the attack.

Studies reported in [6] and [11] indicate that, at the EV layer, PnC mechanisms, certificate management processes, and real-time data exchange can be compromised through HomePlug Green PHY eavesdropping, broken-wire PLC disruption, and certificate or session manipulation arising from weaknesses in ISO 15118 implementations. In addition, the authors in [12] explain that attackers can falsify SoC values and induce denial-of-charging or unsafe charging conditions by manipulating EV-EVSE messages.

Existing research on EV charging cybersecurity provides

valuable but fragmented contributions, as summarized in [1]. Early work in [2] and [5] investigates authentication and certificate-handling mechanisms within ISO 15118, identifying weaknesses in TLS configuration and certificate validation; however, these studies stop short of implementing a complete TLS 1.3 security stack. Complementary analyses in [4] examine vulnerabilities in certificate-based communication, highlighting risks associated with improper key storage and insufficient cryptographic enforcement. Broader threat-modelling efforts (e.g., [5], [6], [7], and [8]) map attack vectors across EV-EVSE communication channels, revealing threats such as spoofing, replay attacks, and supply-chain exploitation. Nevertheless, these works primarily characterize attack surfaces without proposing deployable countermeasures aligned with ISO 15118-20 [9], [24].

Further analyses in [10] and [13] clarify the operational and PKI requirements for PnC but remain largely conceptual, lacking experimentally validated protocol implementations. Research addressing backend security, including [11], [12], and [13], demonstrates that OCPP channels introduce significant risks due to weak authentication flows, insufficient message integrity verification, and vulnerable firmware-update mechanisms. However, these studies consider backend communication largely in isolation from frontend security protocols.

Hardware-based security approaches, such as those presented in [6], propose Trust EV and HSM-based key protection mechanisms, yet their effectiveness is not evaluated under the mandatory mutual TLS requirements introduced in ISO 15118-20. Finally, intelligent security approaches explored in [15] investigate machine-learning-based intrusion detection for vehicular networks, offering promise for EVCS anomaly detection, but without direct consideration of ISO 15118 message sequencing or TLS handshake behaviour.

Table 1 consolidates and synthesizes key contributions from the existing literature on EVCS cybersecurity while highlighting persistent gaps that motivate this research. Rather than providing a detailed comparison of individual studies, the table groups prior work according to core security themes, including transport-layer protection, authentication mechanisms, protocol vulnerabilities, PnC architectures, backend communication security, and data-driven detection approaches. This thematic aggregation enables clearer identification of structural limitations across the field.

Table 1. Selected Literature on TLS with Protocols

REF.	KEY ACHIEVEMENTS	IDENTIFIED GAPS / IMPROVEMENT AREAS
[2], [4], [6], [19], [20]	Definition of TLS-based security requirements; validation of certificates; analysis of TLS handshake, authentication, and encryption mechanisms; recognition of mutually authenticated TLS as a critical security enhancement	Limited analysis of certificate lifecycle management, deployment challenges, and cyber-physical impact of TLS failures on charging outcomes
[5], [3]	Implementation and validation of authentication mechanisms, including user and process authentication; exploration of behavioural biometrics as a second authentication factor	Lack of integration between authentication events and physical charging performance metrics
[6], [18], [24]	Identification of protocol-centric vulnerabilities in ISO 15118-2 communication between EV and	Absence of analysis for ISO 15118-20 upgrades and limited evaluation of

	EVSE due to weak or absent encryption	mitigation effectiveness using modern TLS configurations
[10], [14], [17], [21]	Analysis of Plug-and-Charge (PnC) ecosystem requirements; PKI architectures; trusted platform modules for secure storage of cryptographic material; V2G functional and security requirements	Insufficient evaluation of PnC security under active cyberattack scenarios and limited linkage to operational charging reliability
[8], [13], [12], [15], [7]	Comprehensive threat and vulnerability analyses of EV charging communication protocols; identification of cyberattack vectors affecting EV, EVSE, and charging infrastructure	Predominantly qualitative assessments without causal or quantitative linkage to energy delivery or battery State of Charge degradation
[11], [25]	System-of-systems and layered security approaches for isolating threats across communication layers; distributed infrastructure protection strategies	Limited integration of layered threat isolation with protocol-level security state transitions and charging dynamics
[12], [19], [28], [27]	Security analysis of OCPP-based communication between EVSE and CPOs; identification of vulnerabilities in OCPP profiles; compatibility and security profile verification	Backend security weaknesses insufficiently linked to frontend charging impacts; lack of end-to-end EV–EVSE–CSMS evaluation
[22], [23]	Machine learning applications to improve communication quality and enable intelligent, adaptive secure communication	Limited validation in safety-critical EV charging contexts and weak coupling to protocol standards and physical charging effects

Overall, the literature summarized in Table 1 demonstrates substantial progress in defining cryptographic requirements, analysing protocol vulnerabilities, and proposing layered security architectures for EV charging systems. However, most existing studies remain protocol-centric or largely qualitative, with limited investigation into how cybersecurity failures propagate into physical charging behaviour. Persistent gaps remain in certificate lifecycle management, systematic evaluation of the security enhancements introduced by ISO 15118-20, analysis of backend-to-frontend attack propagation, and understanding the causal relationship between communication-layer disruptions and charging outcomes.

From a cybersecurity perspective, as discussed in [14], the frontend layer involves the EV performing critical functions such as PnC authentication, EVSE discovery, billing data exchange, real-time session authorization, and V2G communication. These functions have been shown to be vulnerable to spoofing, session hijacking, and data manipulation when TLS is absent, optional, or misconfigured (e.g., [1], [18], [5], [19], [20]). Furthermore, work in [5] demonstrates that the EVSE acts as a central orchestration point, responsible for power delivery control, certificate validation, monitoring, and other security-critical operations, often supported by hardware anchors such as HSMs. These hardware-based mechanisms play a key role in protecting cryptographic material and enforcing PKI trust chains, as emphasized in recent studies on secure PnC and EV charging security architectures.

On the backend, the CSMS performs high-value operations including payment processing, user authentication, analytics, grid integration, and firmware updates. Studies such as [16] and [21] identify the CSMS as a prominent attack surface, where

successful compromises can result in financial fraud, privacy violations, remote shutdown of charging fleets, and manipulation of grid-connected charging loads [22], [8].

The bidirectional communication flows illustrated in the figure reflect the continuous exchange of operational and billing data across EV, EVSE, and CSMS components. Consistent with prior findings, these communication paths can be exploited by adversaries to induce functional disruptions, financial and privacy compromises, safety hazards, and even large-scale power system instability [17], [23].

By situating these communication flows within a unified architectural view, the figure emphasizes that EV charging cybersecurity cannot be addressed at a single layer. Instead, effective protection requires coordinated, end-to-end security controls, including mandatory TLS, robust certificate lifecycle governance, secure firmware update mechanisms, and anomaly detection across EV, EVSE, and CSMS components. This perspective aligns with recent studies on cyber-physical risks in EV charging infrastructure and ongoing standards evolution discussed in [5], [24], [25], and [19].

III. VULNERABILITIES IN EVSE

More recent studies (e.g., [9], [28], [26]) extend earlier findings by examining EVSE security from a system-level perspective, demonstrating that vulnerabilities span hardware components, communication protocols, and cloud-connected services. In particular, [29] reinforces concerns regarding the persistence of insecure ISO 15118 deployments and the slow adoption of comprehensive TLS protections.

Building on this, [21] presents a detailed empirical assessment of EVSE attack surfaces, identifying exploitable weaknesses across multiple interfaces, including EV connectors, user authentication terminals, OCPP-based backend services, and maintenance ports. Rather than introducing new attack classes, these results highlight how existing weaknesses manifest simultaneously across interconnected EVSE subsystems.

Experimental evidence reported in [11] [13], [15] and [16] demonstrates how such weaknesses can be exploited in practice through communication-layer interception, PLC disruption, insecure web interfaces, and backend compromise, while [20] further documents firmware-level attacks that enable persistent adversarial control. Collectively, these studies indicate that EVSE security failures are rarely isolated and instead propagate across functional boundaries due to insufficient isolation and inconsistent security enforcement.

Importantly, recent analyses such as [27] and [28] move beyond vulnerability identification to emphasize operational consequences, showing that EVSE compromises can escalate into denial-of-charging events, falsified metering, malware dissemination, and coordinated manipulation of charging loads. These observations reinforce the need for standardized, end-to-end security controls rather than incremental or component-specific defences.

Consequently, the literature increasingly converges on the necessity of enforceable security baselines particularly mandatory TLS, coherent certificate lifecycle governance, and secure update mechanisms providing clear motivation for evaluating the security guarantees introduced by ISO 15118-20 and their effectiveness against emerging cyber-physical threats in EV charging ecosystems.

IV. IMPACTS OF ATTACK ON EVSE CYBERSECURITY

Cybersecurity vulnerabilities in EVSE can give rise to significant functional, financial, safety, and grid-level impacts, underscoring the critical importance of robust protection mechanisms across the EV charging ecosystem. At the functional level, successful cyberattacks may disrupt the availability of individual charging units, entire EVSE fleets, or vendor-managed infrastructures, thereby threatening service continuity as electric mobility expands into safety-critical domains such as emergency response, healthcare, agriculture, and industrial operations. For example, work in [32] demonstrates that malicious remote firmware updates can be leveraged to render charging networks inoperable. Manipulation of communication between the EV and EVSE further exacerbates these operational risks.

Multiple studies [33], [34] report that falsification of SoC information can result in denial-of-charging events, vehicle stranding, and broader disruptions to transportation systems. Moreover, inaccurate SoC reporting may impose excessive stress on battery management systems and increase the likelihood of unsafe charging behaviour in the absence of effective fail-safe controls [34]. Beyond operational impacts, unauthorized access to EVSE and backend systems introduces substantial financial and privacy risks. Prior work in [35] identifies threats including the exfiltration of personally identifiable information, manipulation of billing processes, and compromise of payment credentials. Independent security assessments in [36],[37] further indicate that inadequately secured EVSE deployments can serve as entry points into corporate networks, enabling espionage activities and large-scale data breaches.

Safety-related consequences arise when protective mechanisms within the EV or EVSE are degraded or circumvented. While experiments reported in [34] indicate that internal vehicle protection systems successfully prevented overcharging during a DC fast charger exploitation scenario, compromised communication channels or disabled safety interlocks could undermine these safeguards. Additionally, [38] highlights that CHAdeMO-based charging systems may permit adversarial manipulation of battery operating parameters if IEC 61851 CAN-bus messages are insufficiently validated or filtered.

Emerging charging technologies introduce further safety considerations. High-power liquid-cooled charging cables and wireless power transfer systems, for instance, raise concerns related to electromagnetic interference, including potential impacts on nearby medical devices, as documented in [38]. At the grid level, malicious manipulation of EVSE loads poses systemic risks to power system stability. Experimental investigations in [33] show that attacks targeting EVSE power electronics can induce total harmonic distortion exceeding 20%, reduce power factor below 0.8, and cause abrupt power reductions from 50 kW to 0.3 kW within 0.020 seconds.

Coordinated charging attacks further amplify these risks. Studies in [40] and [41] demonstrate that synchronized malicious charging behaviour can violate voltage constraints, overload network assets, and disrupt economic dispatch processes. At scale, such attacks may trigger under-frequency load shedding events, while [42] suggests that increasing EV penetration could enable adversary-controlled bulk frequency modulation.

EVSE cybersecurity vulnerabilities have the potential to escalate from localized disruptions to widespread functional, safety, financial, and grid-level consequences. These findings

emphasize the necessity of comprehensive, end-to-end cybersecurity strategies to mitigate cascading failures and protect both transportation and power system infrastructures.

V. EMERGING CYBER THREAT IN EVSE INFRASTRUCTURE

Recent literature (e.g., [17], [28], [30],[31]) highlights a rapid increase in cyberattacks targeting critical infrastructure, emphasizing the growing need for coordinated cybersecurity strategies and incident-response frameworks at a global level. Within the EV ecosystem, these threats increasingly affect charging infrastructure and its control backbones, where disruption of communication flows can interfere with OCPP sessions and ISO 15118 authentication processes. Advanced persistent threats have been identified as a major concern, as attackers may infiltrate CSMS environments to compromise authentication credentials, extract sensitive operational data, or manipulate coordinated charging behaviour [43].

Comparable ransomware incidents in other critical sectors, such as healthcare systems [44], illustrate how similar attack strategies could be applied to EV charging networks. In such scenarios, adversaries could encrypt charging management servers, rendering billing systems, session records, and operational dashboards unavailable to EVSE operators. In parallel, the increasing use of distributed denial-of-service (also known as DDoS) attacks based on IoT botnets [45] demonstrates the feasibility of large-scale traffic flooding against EV charging infrastructures. These attacks could overwhelm OCPP communication channels, disrupt authorization services, and cause widespread service outages.

Insights from cybersecurity research on SCADA systems [46] provide further relevant parallels, particularly for EVSE-grid coordination. Manipulation of telemetry data or control signals in these environments has been shown to destabilize load management functions, raising concerns that similar attack techniques could affect EV charging systems governed by power-system security standards such as IEC 62351-7:2017.

Beyond technical attack vectors, broader cybercrime trends introduce additional governance and policy challenges for EV charging ecosystems. Studies on financial cybercrime [47], persistent social-engineering campaigns [48], and inconsistent regulatory enforcement across jurisdictions [49] reveal structural weaknesses in consumer protection, payment security, and identity management within charging services. Research on the behavioural drivers of cybercriminal activity [50], together with documented denial-of-service attacks on public and financial institutions [51]-[53], further illustrates the growing use of service disruption as a strategic objective.

In the context of EV charging networks, such large-scale and coordinated attacks could degrade authentication services, disrupt CSMS operations, and impair backend billing platforms, ultimately undermining service reliability and public trust. These developments reinforce the need for harmonized cybersecurity standards, mandatory resilience requirements, and coordinated incident-response mechanisms to support the secure and dependable deployment of electric mobility infrastructure.

VI. EVSE CYBERATTACKS AND DEFENCES

This work is further extended to address both cyberattacks and corresponding defence mechanisms in EVCS. To provide a general overview, Table 2 summarizes common cyberattack types relevant to EVCS, including denial-of-service, replay attacks, phishing, jamming, GPS spoofing, SQL injection, man-in-the-middle attacks, and electromagnetic interference.

Table 2. Expanded EVCS Cyberattack Types

Attack Type	Brief Description	Potential Impact on EV/EVCS	Ref
Denial-of-Service (DoS)	Overloads the charging system or backend servers with excessive requests, causing service outages.	Blocks access to charging stations, prevents charging, disrupts fleet operations.	[12, 20, 29]
Replay Attack	Captures legitimate communication messages and replays them to trick EV or EVSE.	EV may accept unauthorized commands or false authentication messages.	[11, 30]
Side-Channel Attack	Uses power consumption, EM emissions, or timing leaks to infer private keys.	Extraction of cryptographic keys used in ISO 15118 or OCPP security.	[30, 31]
Phishing Attack	Tricks users into revealing sensitive data or installing malware.	EV owners download malicious apps, firmware, or login credentials are stolen.	[20, 32]
Jamming Attack	Introduces radio interference to disrupt wireless communication channels.	Loss of connectivity in Wi-Fi, Bluetooth, PLC or wireless charging communication.	[16, 33]
GPS Spoofing	Sends falsified GPS signals to mislead navigation or location-dependent charging functions.	Misroutes EVs, affects smart-charging schedules, causes safety risks.	[31, 34]
Man-in-the-Middle (MitM)	Intercepts and manipulates communication between EV and EVSE.	Data theft, session hijacking, injection of harmful commands.	[11, 12, 35]
Buffer Overflow	Sends more data than a system can handle, triggering crashes or malicious code execution.	EVSE shutdown, firmware compromise, remote code execution.	[12, 35]
SQL Injection	Inserts malicious commands into backend database queries.	Tampering with billing data, authentication logs, or firmware configuration.	[12, 19]

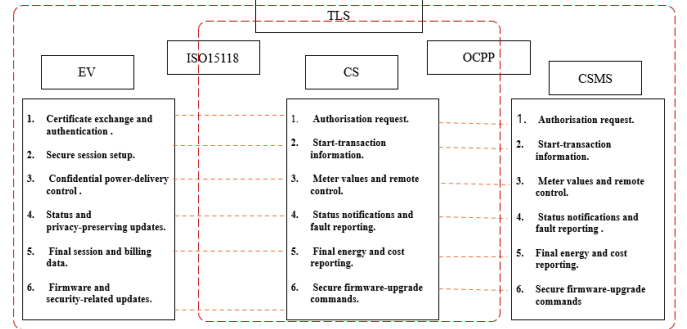
As summarized in Table 2, cyberattacks on EVCS can affect confidentiality, integrity, availability, and safety across both EV and EVSE environments. The classification highlights a wide range of attack vectors, including protocol-level manipulation, backend exploitation, and human-focused attacks. This diversity of threats demonstrates that isolated or single-layer defences are insufficient and reinforces the need for comprehensive security strategies, as noted in [28] and [54].

Recent studies (e.g., [17], [10], [36], [56]) show that cyberattacks targeting EVCS are becoming increasingly sophisticated and have broader system-level consequences. Security analyses of CSMS platforms and OCPP-based backend systems reveal that attackers commonly exploit protocol weaknesses, inadequate authentication mechanisms, and insecure application programming interfaces. These vulnerabilities enable attacks such as data theft, billing manipulation, remote disabling of chargers, and large-scale service disruption.

Beyond backend systems, research in cyber-physical power systems indicates that compromised charging behaviour can propagate beyond individual charging stations. Studies in [37], [37], [39], and [41] demonstrate that coordinated load-altering attacks such as botnet-driven manipulation of charging behaviour and false data injection in EV-grid communication can destabilize distribution networks, disrupt voltage and frequency regulation, and reduce overall system reliability. These effects are further amplified under conditions of high vehicle-to-grid penetration.

One may conclude that cybersecurity risks in EVCS are inherently multi-layered, spanning physical infrastructure, communication protocols, and cloud-connected backend services. This leads to the consideration and the implementation of effective protection that may require coordinated and systematically applied security controls across all architectural layers, mainly to prevent cascading failures and mitigate operational, safety, and grid-level impacts.

In response to these challenges, several mitigation approaches are currently under investigation. Among them, the implementation of TLS is widely considered an effective means of reducing the risk of cyberattacks on EVCS communications. Figure 2 illustrates the role of TLS within the overall EVSE communication architecture. Specifically, the figure presents the end-to-end communication workflow of an EVCS, highlighting the interactions between the EV, the CS, and the CSMS under the ISO 15118 and OCPP standards.

**Fig. 2:** End-to-End TLS-Secured Message Exchange Between EV, EVSE, and CSMS.

As can be seen in this figure, the end-to-end communication workflow among the EV, EVSE, and CSMS, depicting the structured exchange of messages throughout a charging session. The figure outlines key operational phases, including certificate-based authentication, secure session establishment, real-time control signalling, metering, billing, status reporting, and firmware updates, thereby contextualizing previously identified threat surfaces within a unified system architecture. It further demonstrates how the combined use of TLS 1.3 with ISO 15118 and OCPP provides security coverage across both frontend and backend communication paths. The mandatory mutual authentication and enhanced certificate management mechanisms introduced in ISO 15118-20 strengthen early authentication and session establishment stages that have been shown to be vulnerable to replay, impersonation, and man-in-the-middle attacks, while the encrypted and structured message exchanges defined in OCPP 2.0.1 mitigate backend threats such as unauthorized control commands, manipulation of metering data, and unauthorized access to billing or operational information. By presenting TLS as a common security layer spanning both protocol domains, the figure provides a systematic basis for evaluating how recent standard developments reduce attack surfaces and improve end-to-end security in EV charging infrastructures.

VII. ANALYSES AND DISCUSSION

The findings presented in this study confirm that cybersecurity in EVCS remains a critical and largely unresolved challenge, driven by fragmented protocol adoption, inconsistent enforcement of security controls, and the increasing sophistication of adversarial capabilities. As noted in the Introduction, the persistent lack of TLS deployment reported in more than 80% of operational charging stations continues to expose ISO 15118-2 and OCPP-based communication channels to attacks that compromise the confidentiality, integrity, and availability of charging services. This concern is consistently reinforced across the literature (e.g., [1]-[28]), which documents vulnerabilities in ISO 15118 message flows, misconfigured PKI infrastructures, optional or outdated encryption mechanisms, and weak backend authentication practices that collectively undermine the security posture of contemporary EVCS deployments.

A recurring limitation identified in prior work (e.g., [2], [9]–[6]) is the architectural fragmentation between frontend and backend security analyses. Frontend-focused studies primarily investigate vulnerabilities in EV–EVSE communication, such as spoofing, replay attacks, session manipulation, and PLC interception within ISO 15118, but often neglect dependencies on backend systems. Conversely, backend-oriented research (e.g., [8], [12], [10]) emphasizes exploitable APIs, weak authentication flows, insecure firmware update mechanisms, and SQL injection vulnerabilities within CSMS platforms, typically without integrating frontend protocol security considerations. This separation reveals a significant research gap, as limited work evaluates how ISO 15118-20’s mandatory TLS 1.3 interacts with OCPP 2.0.1 security mechanisms within a complete, end-to-end charging workflow.

This gap is particularly consequential given that adversaries frequently exploit cross-layer dependencies. As shown in [19] and [20], compromised EVSE firmware or man-in-the-middle positioning can facilitate escalation into backend systems, while backend breaches can enable manipulation of authorization, billing, or power-delivery parameters across entire charging fleets. Beyond technical vulnerabilities, the literature clearly demonstrates the cyber-physical consequences of EVCS attacks. Studies in [24]–[29] show that EVSE compromise can disable individual chargers or entire fleets, falsify SoC values, propagate malware through firmware updates, and disrupt critical services reliant on EV mobility. These findings substantiate earlier concerns that insecure ISO 15118-2 and OCPP deployments may enable denial-of-charging events, unsafe charging conditions, inaccurate billing, and widespread service outages, alongside substantial financial and privacy risks.

Real-world assessments, including those reported in [37], further demonstrate that insecure EVSE deployments can act as entry points into corporate networks, enabling advanced intrusions and large-scale data breaches. At the grid level, studies in [29]–[34] indicate that cyberattacks targeting charging infrastructure can propagate beyond individual devices, threatening both distribution and transmission system stability. Compromised EVSE behaviour (e.g., coordinated load drops, rapid power transitions, falsified metering, and manipulated V2G dispatch) has been shown to induce voltage instability, overload network assets, disrupt generator scheduling, and trigger under-frequency load shedding events. These observations, reinforced by [51] and [52], highlight that EVCS cybersecurity risks are intrinsically linked to broader power system resilience, particularly under conditions of increasing EV penetration and expanded V2G participation.

The diversity of attack vectors summarized in Table 2 (e.g., [1], [18], [22], [43] [53][54][55]), ranging from conventional IT threats such as SQL injection and buffer overflows to cyber-physical attacks including GPS spoofing, PLC jamming, and electromagnetic interference, further illustrates the hybrid threat landscape facing EV charging systems. As emphasized in [28] and [44], the coexistence of these threat classes necessitates defence strategies capable of providing simultaneous protection across hardware, software, communication layers, and human–machine interfaces. Recent studies also indicate that these attacks are becoming increasingly automated, scalable, and adaptive, underscoring the importance of proactive detection mechanisms, secure firmware governance, rigorous PKI lifecycle management, and enhanced operational awareness among EVSE operators.

Finally, the layered architecture illustrated in Figure 1 reinforces these insights by highlighting the tight

interdependencies among EVs, EVSE, and CSMS components. Vulnerabilities at any layer—such as misconfigured TLS at the EV, inadequate certificate validation at the EVSE, or insecure APIs within the CSMS—can be leveraged to compromise the integrity of the entire charging workflow. In this context, the role of hardware security modules at charging stations, as discussed in [19] and [37], is increasingly critical for protecting cryptographic keys and enforcing secure Plug-and-Charge trust chains. Similarly, the backend remains a high-value target due to its responsibility for authentication, billing, grid integration, analytics, and firmware management, as highlighted in [37] and [39]. Taken together, the findings of this work indicate that the security of electric vehicle charging ecosystems depends on the coherent and systematic enforcement of end-to-end safeguards. These safeguards include mandatory TLS 1.3, comprehensive certificate lifecycle governance, tightly controlled firmware management practices, and continuous anomaly monitoring, applied consistently across EV, EVSE, and CSMS components in alignment with evolving cybersecurity standards [39]–[42].

VIII. CONCLUSION

The work has shown that electric vehicle charging systems remain exposed to cyber-physical risks due to fragmented protocol adoption and inconsistent security implementation in legacy ISO 15118-2 among other factors such as OCPP deployments. The review demonstrates that ISO 15118-20 can provide substantial security improvements through mandatory mutually authenticated TLS 1.3 and strengthened public key infrastructure requirements, enabling effective mitigation of replay, impersonation, and man-in-the-middle exploitation attacks when deployed end-to-end. The findings further confirm that electric vehicle charging systems security is inherently cyber-physical, with communication failures capable of escalating from local charging disruption to wider system impacts. Secure electric mobility therefore requires more advanced techniques such as coordinated, system-wide protection across communication protocols, cryptographic key management, firmware integrity, and backend services.

Among many other work, incorporating quantitative cyber-risk assessment approaches, such as probabilistic attack modelling, can help with evaluating the likelihood and impact of attacks on EV charging infrastructure and support more effective mitigation planning.

REFERENCES

- [1] M. Szakály, S. Köhler, and I. Martinovic, “Current affairs: A measurement study of deployment and security trends in EV charging infrastructure,” arXiv preprint arXiv:2404.06635, 2024.
- [2] K. Ahmet, “Smart, secure and interoperable charging infrastructure with Plug and Charge,” in Proc. 12th Int. Conf. Smart Grid (icSmartGrid), 2024.
- [3] J. Sturges et al., “CableAuth: A biometric second-factor authentication scheme for electric vehicle charging,” Tech. Rep., 2023.
- [4] S. Acharya, Y. Dvorkin, and R. Karri, “Public plug-in electric vehicles + grid data: A cyberattack vector?” IEEE Trans. Smart Grid, vol. 11, no. 6, pp. 5099–5113, 2020.
- [5] A. Kilic, “TLS handshake for Plug and Charge in vehicular communications,” Comput. Netw., vol. 243, p. 110281, 2024.
- [6] P. Van Aubel and E. Poll, “Security of EV-charging protocols,” arXiv preprint arXiv:2202.04631, 2022.
- [7] A. Ahalawat, S. Adepu, and J. Gardiner, “Security threats in electric vehicle charging,” in Proc. IEEE SmartGridComm, 2022.
- [8] A. Kilic, “Plug and Charge solutions with vehicle-to-grid communication,” Elect. Power Compon. Syst., vol. 51, no. 16, pp. 1786–1814, 2023.
- [9] J. H. Do, J. K. Ho, and L. D. Hoon, “A study on securing communication of electric vehicle charging using combined charging system,” in Proc. KICS Conf., pp. 198–200, 2021.
- [10] A. Heinrich and R. Heddergott, “Secure and user-friendly EV charging: A comparison of Autocharge and ISO 15118 Plug & Charge,” Hubject GmbH, Tech. Rep., Jun. 2019.

- [11] J. Johnson et al., "Review of electric vehicle charger cybersecurity vulnerabilities, potential impacts, and defenses," *Energies*, vol. 15, no. 11, p. 3931, 2022.
- [12] Z. Garofalaki et al., "Electric vehicle charging: A survey on the security issues and challenges of the Open Charge Point Protocol (OCPP)," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1504–1533, 2022.
- [13] R. Baker and I. Martinovic, "Losing the car keys: Wireless insecurity in EV charging," in *Proc. 28th USENIX Security Symp.*, 2019.
- [14] A. Fuchs et al., "HIP: HSM-based identities for Plug-and-Charge," in *Proc. 15th Int. Conf. Availability, Reliability and Security*, 2020.
- [15] Z. El-Rewini et al., "Cybersecurity challenges in vehicular communications," *Veh. Commun.*, vol. 23, p. 100214, 2020.
- [16] J. Antoun et al., "A detailed security assessment of the EV charging ecosystem," *IEEE Netw.*, vol. 34, no. 3, pp. 200–207, 2020.
- [17] A. Kilic, "Secure and manipulation-proof TLS communication with Plug and Charge," *Tech. Rep.*, 2023.
- [18] S. Hamdare et al., "Cybersecurity risk analysis of electric vehicle charging stations," *Sensors*, vol. 23, no. 15, p. 6716, 2023.
- [19] K. Sarieddine et al., "Uncovering covert attacks on EV charging infrastructure," in *Proc. ACM Asia CCS*, 2024.
- [20] H. Jahangir et al., "Charge manipulation attacks against smart EV charging stations," *IEEE Trans. Smart Grid*, 2024.
- [21] F. Haidar, "Vehicle-to-grid: Towards a cyber-secure electric system including vehicle, charging points and batteries," in *Proc. SIA Powertrain & Energy Congr.*, 2022.
- [22] E. S. Ali et al., "Machine learning technologies for secure vehicular communication in Internet of Vehicles," *Security Commun. Netw.*, 2021.
- [23] K. N. Qureshi, A. Alhudhaif, and G. Jeon, "EV energy management and charging scheduling in sustainable cities," *Sustain. Cities Soc.*, vol. 71, p. 102990, 2021.
- [24] H. El-Hussini et al., "A tale of two entities: Contextualizing the security of electric vehicle charging stations on the power grid," *ACM Trans. Internet Things*, vol. 2, no. 2, 2021.
- [25] R. M. Pratt and T. E. Carroll, "Vehicle charging infrastructure security," in *Proc. IEEE ICCE*, 2019.
- [26] T. Lauser, D. Zelle, and C. Krauß, "Security analysis of automotive protocols," in *Proc. ACM Comput. Sci. Cars Symp.*, 2020.
- [27] R. Metere et al., "Cyber security and privacy in EV charging infrastructure," *arXiv preprint arXiv:2209.07842*, 2022.
- [28] L. Buschlinger et al., "Plug-and-patch: Secure value-added services for EV charging," in *Proc. 14th ARES*, 2019.
- [29] M. Kaiser, "Automating compatibility testing for integration of charge points into an EV ecosystem," *Tech. Rep.*, 2024.
- [30] A. Z. Galbis et al., "Smart tool development for customised EV charging services," *World Electr. Veh. J.*, vol. 13, no. 8, p. 145, 2022.
- [31] P. Aji et al., "Development of EV charging station management systems," in *Proc. IEEE ICT-PEP*, 2020.
- [32] S. Hsaini et al., "An OCPP-based approach for EV charging management," *Energies*, vol. 15, no. 18, p. 6735, 2022.
- [33] Open Charge Alliance, *Open Charge Point Protocol (OCPP) 2.0.1 Specification*, Dec. 2024.
- [34] D. Priyasta et al., "Ensuring compliance of OCPP 1.6 messages," *J. Eur. Syst. Autom.*, vol. 56, no. 1, 2023.
- [35] M. Mülten, "New features and timeline for ISO 15118-20," *Switch-EV*, 2020.
- [36] M. Shin et al., "Interoperability testing for EV chargers," *Appl. Sci.*, vol. 6, no. 6, p. 165, 2016.
- [37] A. Pacheco et al., "Energy transition on geographic islands," *Renew. Energy*, vol. 184, pp. 700–711, 2022.
- [38] Z. Pourmirza and S. Walker, "EV charging station cybersecurity challenges," in *Proc. IEEE SEGE*, 2021.
- [39] R. Flocea et al., "EV smart charging reservation algorithm," *Sensors*, vol. 22, no. 8, p. 2834, 2022.
- [40] S. B. Mitikiri et al., "Cyber-physical security in EV charging infrastructure," *J. Cleaner Prod.*, vol. 438, p. 140347, 2025.
- [41] H. R. Sayarshad et al., "Cyberattack resilience and vehicle-to-grid integration," *Transport Policy*, vol. 139, pp. 158–169, 2025.
- [42] S. Hamdare et al., "Cyber security for EV smart charging systems," *Conf. Paper*, 2025.
- [43] H. Tanyıldız et al., "Detection of cyber attacks in EV charging," *Frontiers Energy Res.*, vol. 13, 2025.
- [44] S. B. Mitikiri et al., "Addressing vulnerabilities in the EV ecosystem," in *Proc. EVS38*, 2025.
- [45] Upstream Security, *Automotive & Smart Mobility Global Cybersecurity Report 2025*, 2025.
- [46] J. Han et al., "Transfer learning for securing EV charging infrastructure," *Sci. Rep.*, vol. 15, 2025.
- [47] A. Wahab et al., "Grid Sentinel anomaly detection framework," *Comput. Intell. Neurosci.*, 2025.
- [48] X. Luo et al., "AI-enhanced intrusion detection for EV charging," *Alexandria Eng. J.*, vol. 74, no. 2, 2025.
- [49] S. Chia et al., "Black-box fuzzing of EV charging protocols," *Comput. Mater. Continua*, vol. 84, no. 2, 2025.
- [50] R. V. Gottumukkala et al., "Cybersecurity in onboard EV charging systems," *World J. Adv. Res. Rev.*, 2020.
- [51] J. AlFarra et al., "ML-driven distributed intrusion detection for EV charging," *CEUR Workshop Proc.*, vol. 3935, 2024.
- [52] S. B. Mitikiri et al., "Anomaly detection of cyber attacks on EV charging ports," *Electr. Power Syst. Res.*, vol. 224, p. 109611, 2024.
- [53] A. Wahab et al., "Machine learning empowered anomaly detection," *Comput. Intell. Neurosci.*, 2024.
- [54] Open Charge Alliance, *OCPP Information and Whitepapers*, 2024.
- [55] "Cyber defense in OCPP for EV charging security risks," *Int. J. Inf. Security*, 2024.

Tech Diplomacy: A Comprehensive Framework for Addressing Critical Technologies

Kamel Dine Remili, Adel Belouchrani, Riad Hartani and Nadjoua Bouzourine

Abstract—This paper discusses the problem of addressing critical technologies through tech diplomacy. The latter has recently emerged as a strategic response to manage the geopolitical consequences that are induced by rivalry over the control of critical technologies. Despite the growing importance of these technologies in global politics, the relationship between technological competition and diplomatic governance remains insufficiently conceptualized. The existing research often lacks an integrated framework connecting critical technologies to geopolitical competition, diplomatic mechanisms, and global governance. Moreover, most analyses focus on state competition while the influence of private technology multinationals remains underexplored despite their major global role. To address this gap, we propose a comprehensive Tech Diplomacy Framework (TDF) which addresses the diplomatic dynamics in international relations that these technologies generate. The framework is subsequently applied to Low Earth Orbit satellite internet as an illustrative critical technology to validate the proposed framework and its applicability to the whole range of critical technologies.

Keywords—tech diplomacy, critical technology, tech diplomacy framework, private tech firms, satellite internet.

I. INTRODUCTION

Technology has long been an important source of power, and a country's technological capabilities are widely understood to influence its economic competitiveness, political influence, and military strength. Historically, technological development was in many cases closely associated with the state. Major technological advances, including nuclear energy, aerospace systems, telecommunications, navigation technologies, and early computing, were often directly developed or substantially supported through public funding. These technologies supported a range of objectives including territorial governance, defence capabilities, administrative capacity, and economic development. While private firms also contributed, the state generally played a central role in setting technological priorities and providing strategic direction. Over time, particularly toward the end of the 20th century, globalization and the emergence of the internet contributed to a gradual transformation of the innovation system in many countries. Governments increasingly shifted away from direct technology development and moved toward supporting market-driven innovation and public-private partnerships [1], [2].

In the 21st century, technologies such as artificial intelligence (AI), semiconductors, cloud and quantum computing, biotechnology, nanotechnology, satellite communications, and robotics are reshaping every domain of human life and affecting economic structures, geopolitical dynamics, and social systems.

Advanced technologies now underpin military readiness, intelligence capabilities, and economic resilience. Yet, unlike earlier strategic technologies, they are overwhelmingly developed and owned by private multinational firms operating in global markets [3]. While states continue to define national priorities, they increasingly depend on privately controlled technologies to implement them. Consequently, private tech firms have become influential non-state actors in global affairs; and states face a reconfiguration of authority and must negotiate and partner with private firms that control critical capabilities. This reality creates new dependencies, redistributes power, and compels states to develop new governance strategies [4].

In this context, diplomacy becomes a tool not only for managing interstate relations but also for engaging the private technology industry in the interest of national security. This situation induces a shift from traditional diplomacy, focused on treaties and state-to-state negotiations, to a new paradigm capable of addressing fast-moving, technically complex, and privately driven innovation ecosystems. This transformation has given rise to new diplomatic practices aimed at coordinating standards, managing technological competition, and fostering cooperation across states, private entities, and international bodies to secure access to critical technologies and shape norms as well as mitigate vulnerabilities. This emerging field, often referred to as technology diplomacy, has become a key mechanism through which states seek to influence the development, diffusion, and governance of critical technologies. Despite the growing importance of technology diplomacy, the relationship between technological competition and diplomatic governance remains insufficiently conceptualized. The existing research often lacks an integrated framework connecting critical technologies to geopolitical competition, diplomatic mechanisms, and global governance. Most analyses focus on state competition (such as US-China rivalry) while the influence of private technology firms remains underexplored, despite their major role in sectors like AI, semiconductors, space telecommunications, and digital infrastructure.

This paper proposes a comprehensive framework for addressing critical technology through tech diplomacy. It first starts by defining the concepts of critical technologies and tech diplomacy. Then the proposed Tech Diplomacy Framework (TDF) is described. As an example of application, the proposed

Manuscript received April 5, 2026; accepted June 2, 2026.

K. D. Remili and A. Belouchrani are with Electronics Department, Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: kamel_dine.remili@g.enp.edu.dz, adel.belouchrani@g.enp.edu.dz).

R. Hartaniis with Xona Partners, USA (e-mail: riad@xonapartners.com).

N. Bouzourine is with Department of Public Policies and Comparative Systems, Ecole Nationale Supérieure De Sciences Politiques (e-mail: bouzourine.nadjoua@enssp.dz).

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.364

framework is applied to Low Earth Orbit (LEO) satellite internet which is viewed not as “just space” or “just telecoms” nor as “just internet governance”, but as a geopolitical infrastructure that requires dedicated diplomatic capacity.

II. CONCEPTS AND DEFINITIONS

In this section, we briefly provide the concepts and definitions of critical technologies and Tech Diplomacy.

A. Critical Technologies

A practical approach to comprehend the term “Critical Technologies” (CTs) is to refer to countries' national strategy policies [5]. They are commonly described as advanced technologies that are most significant to achieving national security and are generally termed in ways related to sovereignty and prosperity's are not simply economic assets but strategic resources that are essential for national capabilities and whose control or disruption could significantly affect national interests.

Typically, countries define their CTs on the ground of [6]:

- *Importance to national security and defense:* agencies responsible for defense and intelligence provide input on technologies that need to be protected from foreign threats like espionage, intellectual property theft, and sabotage in key sectors such as energy, communications, finance, and defense;
- *Impact on economic and industrial competitiveness:* governments assess which technologies are vital for maintaining and enhancing their country's economic position. This includes evaluating sectors that contribute to Gross Domestic Product (GDP) growth, job creation, and global competitiveness;
- *Supply chain resilience:* countries consider vulnerability in their supply chains. If technology is reliant on foreign suppliers, especially those from adversarial nations, it is more likely to be considered critical;
- *Global trends and geopolitical considerations:* the CTs' lists are constantly updated in response to new threats, technological advancements, and geopolitical changes. In an era of U.S-China competition, many countries reassess technologies such as AI, advanced semiconductors, and biotechnology to ensure their independence from foreign powers.

CTs often involve cross-border supply chains, international standards, and technology alliances. Countries may impose export controls, investment restrictions, or data protection laws. Therefore, CTs' governance extends beyond domestic policy and enters the realm of diplomacy.

B. Tech Diplomacy

Tech Diplomacy emerged in the 21st century digital era, driven by the geopolitical influence of CTs, technology firms, and digital infrastructures. Its core purpose is to manage technological power relations between states, corporations, and international institutions to shape technology norms and governance. The primary actors include nation-states, private sector technology companies, norms and standards-setting bodies, multi-stakeholder coalitions (such as the UN Tech Envoy and the Global Partnership on AI), academia, and civil society. Its institutional model is based on networked diplomatic architecture operating across states and the private sector. Its

main diplomatic logic is based on the recognition that technology embodies geopolitical, ethical, and economic power. Tech diplomacy promotes influence through technological standards, data sovereignty, and leadership in digital governance. Its primary arenas are tech hubs (e.g. Silicon Valley (USA), Shenzhen (China), and Bengaluru (India)); digital summits (e.g., GovTech and India-AI Impact), and corporate-state negotiation spaces. Its temporal orientation concerns rapid adaptation to technological disruption, innovation cycles, and global crises. The underlying Philosophy under tech diplomacy is “*Technology for governance and power balance.*” In summary, we define tech diplomacy as:

“A technology oriented statecraft to engage governments, international bodies, civil society, and tech industry –spanning the full range of technologies- in order to proactively shape norms, governance and power relations in pursuit of national interests and technological sovereignty.”

A distinctive feature of tech diplomacy is the centrality of private technology firms as strategic actors. As today's critical capabilities are predominantly owned, developed, and operated by private corporations, these firms exercise significant autonomy in shaping technological trajectories, setting industry standards, controlling chokepoints in global supply chains, and making deployment decisions with direct geopolitical consequences [7]. This reconfiguration raises fundamental questions about the distribution of sovereignty in the international order [3] and requires states to develop diplomatic capacities oriented towards corporations that increasingly function as quasi-sovereign actors in the global technology order [8]. Tech diplomacy must therefore engage private firms not simply as regulated entities or procurement partners, but as actors whose decisions (on where to deploy infrastructure, which governments to serve, which standards to adopt, and how to respond to state coercion) constitute acts of political consequence.

C. Related Work and Research Gap

The scholarly literature on technology and international relations has grown substantially, however it remains fragmented across three largely parallel streams: power and statecraft theory [7], [9], interdependence and economic coercion [4], and multi-stakeholder governance [10] [11].

Taken together, these streams illuminate different facets of the technology-diplomacy nexus: power dynamics, structural dependencies, and governance contestation. However, no existing framework integrates these dimensions into a coherent, actionable architecture applicable across different critical technology domains. Moreover, private technology firms, despite their central role in shaping the development, deployment, and governance of critical technologies, remain largely peripheral in both the statecraft and governance literatures. The Tech Diplomacy Framework proposed in Section III aims to fill this gap by offering an integrated, layered model that connects strategic framing, supply chain politics, governance norm-setting, and security management, while explicitly accounting for the growing agency of private sector actors.

III. PROPOSED TECH DIPLOMACY FRAMEWORK FOR ADDRESSING CRITICAL TECHNOLOGY

In this section, we propose the Tech Diplomacy Framework (TDF), a comprehensive framework that addresses any critical technology through tech diplomacy. This framework

conceptualizes state engagement with CTs as a layered process that unfolds across four interdependent dimensions: strategic identification and framing, infrastructure and supply chain engagement, technical governance and standards, and security, crisis, and alliance management. It specifies how states should mobilize diplomatic tools across each dimension and offers a cross-domain architecture that connects material, normative, and strategic dimensions into a coherent whole.

This TDF is not tied to specific sectors but instead captures recurring patterns through which states seek to govern, contest, and leverage technological power. It also stresses the role of private technology firms, international bodies, and civil society as actors that states must engage through diplomatic means.

Finally, the proposed framework positions tech diplomacy as a core practice in contemporary international relations (IR) rather than considering it as technical adjunct to foreign policies.

A. Strategic Identification and Framing

The first layer identifies the technologies considered strategically important. Technologies are not designated as critical solely based on their technical characteristics; they are identified as such through sovereignty assessment, economic prioritization, and strategic foresight. Governments classify technologies as critical based on their perceived domestic impacts and strategic geopolitical relevance.

Once specific technologies are deemed critical, policymakers construct narratives linking these CTs to national security and sovereignty, economic competitiveness, and societal resilience, thereby legitimizing extraordinary policy attention and diplomatic engagement. This framing is conveyed through national strategies and foreign policy.

At this stage, tech diplomacy's objective is to understand technological trends and power distribution trajectories by monitoring the global innovation system and assessing technological dependencies through technology foresight programs, science and technology networks, and technology monitoring units in foreign affairs ministries.

B. Infrastructure and Supply-Chain Engagement

CTs' production often involves distributed international value chains that create strategic vulnerabilities in maintaining stable access to technology inputs. Once a technology is framed as critical, tech diplomacy efforts move to the material layer of that technology, where the production networks, logistics systems, physical infrastructures, and transnational supply chains that underpin that technology become central.

Because most advanced technologies rely on globally distributed supply chains, governments must use diplomatic tools to secure access to components, manufacturing capacity, energy, raw materials, technological infrastructures and Intellectual Property. Typically, these resources are unevenly distributed, difficult to substitute in the short term, and often subject to changing policies; as a result, they create asymmetric dependencies that benefit their owners and contribute to what Farrell and Newman describe as infrastructure power [4].

In this layer, tech diplomacy often intersects with industrial policy and economic statecraft, reflecting the blurred boundary between foreign and domestic economic governance. Here, tech diplomacy aims to secure a resilient supply chain and sustain the sourcing of critical inputs (such as raw materials, equipment, and performing hardware). Diplomatic engagement

targets allies and suppliers, as well as private firms that control chokepoints in the value chains.

Tech diplomacy actions encompass a range of initiatives, including supply chain diversification policies, supply chain agreements, strategic partnerships, joint manufacturing initiatives, as well as establishing trusted technological partnerships, participating in the technological supply-chain by building regional manufacturing hubs, investing in shared infrastructure projects, and structuring the financial models to access such technologies.

C. Technical Governance and Standards

While standards, protocols, and design choices determine interoperability, scalability, security, and long-term technological trajectories, global technology governance becomes a strategic, diplomatic, and societal imperative [12]. Effective governance aims to align technical and political objectives and turn national interests into interoperable rules.

Here, tech diplomacy operates through sustained engagement with international standards bodies, expert communities, academia, and industry consortia. States deploy diplomats, regulators, engineers, and corporate actors to influence agendas and outcomes, seeking to embed their preferred norms and standards into global systems. This means coordinating across different administrations domestically and participating in global technology governance initiatives, data governance regimes, digital trade agreements, cybersecurity rules and ethical frameworks for emerging technologies.

This layer is particularly significant because technical decisions often precede and constrain later regulatory or political choices, locking advantages for early movers and indirectly shaping global governance. States may seek to export their regulatory models, resist external rules, or construct coalitions around shared governance approaches.

D. Security, Crisis, and Alliance Management

At the strategic layer, critical technologies are integrated into security planning and alliance strategies. Technologies are treated as dual-use capabilities, force multipliers, or vulnerabilities, leading to coordination on protection, denial, or joint development. Tech diplomacy here is often inward-looking toward trusted partners, focusing on alliance cohesion, risk management, and escalation control.

This layer also highlights the growing role of private technology firms as strategic actors whose decisions can affect national security outcomes. Managing public-private interdependence becomes a diplomatic task, reinforcing the need for continuous engagement beyond formal interstate negotiations.

The structure of the proposed technology-agnostic framework allows it to be applied across various domains, such as artificial intelligence and biotechnology to energy systems and space infrastructure. It offers a coherent analytical perspective for understanding how states seek to consolidate their position in a technologically driven international order, in which critical technologies are embedded in sovereignty, development, and privacy.

Taken together, the four layers constitute an integrated diplomatic architecture rather than a sequential checklist. A state's position in any one-layer shapes and constrains its options in the others: a country that fails to frame a technology as strategically critical (Layer 1) will under-invest in supply chain diplomacy (Layer 2), arrive at governance negotiations

without a coherent position (Layer 3), and find itself without pre-negotiated arrangements when a crisis emerges (Layer 4). By contrast, early and thoughtful engagement across all layers can increase advantages. This interdependence is the framework's central analytical claim: tech diplomacy is most effective not when deployed reactively in a single dimension, but when pursued as a coordinated, cross-layer strategy.

A final and cross-cutting dimension of the framework concerns the role of private technology firms. The TDF explicitly positions private firms as co-constitutive actors whose decisions shape outcomes across all four layers. In the strategic framing layer, firms co-produce the narratives and threat assessments that drive governmental classification decisions. In the supply chain layer, corporations controlling critical chokepoints exercise genuine bargaining leverage. In the governance layer, firms engage standards bodies, shape technical architectures, and build coalitions independently, often outpacing governmental processes. Finally, in the security layer, as Starlink's deployment during the war in Ukraine illustrates, private satellite operators can function as *de facto* strategic actors, making access, coverage, and prioritization decisions with direct military consequences that no existing interstate treaty governs [8]. Tech diplomacy must therefore be understood not only as a state-to-state or state-to-institution practice, but also as a triangular engagement among states, international bodies, and private firms whose structural power increasingly rivals that of sovereign actors in the global technology order [3], [7].

IV. APPLICATION: TECH DIPLOMACY APPLIED TO SPACE INTERNET

In this section, we apply our proposed framework to address Low Earth Orbit (LEO) satellite internet through Tech diplomacy.

A. Context

As the LEO satellite ecosystem expands, governments must adopt a strategic, security-conscious approach to partnerships. Although services like LEO satellite internet offer transformative potential, they also raise political, economic, and sovereignty risks, such as surveillance and external control. To safeguard digital sovereignty, nations should seek balanced partnerships that ensures equitable, long-term collaboration. These objectives are carried out by tech diplomats, a new breed of IR practitioners.

Addressing satellite internet means addressing the multifaceted legal landscape surrounding this technology [13]. Aspects like services and frequencies licensing (varying from one country to another and presenting diverse requirements for satellite internet providers), cross-border service provisions (involving intricate international laws and agreements¹) and liability issues (in cases of satellite malfunctions or collisions for instance) call for robust legal frameworks and extensive international cooperation to address these challenges effectively [6]. Moreover, they underscore the necessity for constant dialogue and collaboration between industry players, regulatory bodies, and governments.

Tech diplomats need to participate in developing new international space agreements and work on existing ones as well as establishing norms for good space behaviour. They ought to get involved in regulating access to space and ensure satellite internet activity does not disturb existing scientific and

industrial activities like observing comets and meteorites. They are required to address data privacy and security risks, as well as to monitor the dual use of that technology and its militarization due to the growing strategic value of satellite internet for defence and security.

From a technical stance, tech diplomats need to be adequately tooled to comprehend all the implications that result from the adoption of new standards. Furthermore, there is a necessity to address the LEO constellations' environmental impact (such as light pollution and carbon footprint), the constellation protection and disaster recovery, the Critical Space Infrastructure Protection (CSIP), as well as the frequency allocation processes within and across countries. Tech diplomats need to create synergy amongst designers, investors, regulators, manufacturers, and operators to raise awareness on the different impacts [14].

B. TDF applied to LEO Satellite Internet

LEO satellite internet is not “just space,” or “just telecoms,” nor “just internet governance”, but it is a geopolitical infrastructure that requires a dedicated diplomatic approach, as illustrated through the use of our proposed Tech Diplomacy Framework:

Layer 1: Strategic Identification and Framing

Based on its previously addressed challenges and geopolitical aspects, this layer frames LEO satellite internet as a critical technology and a strategic infrastructure, rather than just as a commercial service. Governments should position satellite internet as vital for national resilience, defence readiness, and economic sovereignty. They also need to integrate it into foreign, defence, and development policies while assessing the risks associated with reliance on foreign operators. This translates into:

- Integrating LEO satellite internet into narratives about national security and resilience;
- Reclassifying satellite connectivity as critical or dual-use infrastructure to enable governance across ministries and exceptional regulatory authority;
- Conducting strategic risk assessments related to dependency on foreign-controlled orbital systems, focusing on vulnerability and resilience.

If countries choose not to consider LEO satellite internet as inherently linked to their security, sovereignty, and strategic control, they still need to assess the risks and consequences of this technology, which will affect them directly (due to constellations overflying, space congestion, and potential unauthorized rogue use of the services for instance) as well as indirectly if neighbouring countries use LEO satellite services. Furthermore, they need to consider the governance layer outlined below.

Layer 2: Infrastructure and Supply-Chain Engagement

This layer focuses on the long-term structural dependency of the LEO satellite internet. Unlike terrestrial networks, LEO systems allow foreign providers to deliver essential connectivity within a country without necessarily building or owning significant infrastructure there. This increases the risk of external control over domestic connectivity, which means that decisions about access, pricing, data governance and service continuity may occur beyond national control. Rather than simply facilitating market entry, tech diplomacy emerges here as a mechanism for politically negotiating development outcomes and mitigating

¹Such as ITU Radio Regulations, WTO and its General Agreement on Trade in Services (GATS) commitments, and Bilateral and Multilateral Agreements.

dependency issues between global satellite operators and host states. The goal is to secure the benefits of LEO satellite internet while avoiding new forms of technological dependency, regulatory weakening, or uneven market power. Diplomatic actions could include negotiating pricing, service coverage and local partnerships while leveraging regional diplomacy and embedding satellite internet in national digital strategies. These objectives can be achieved by:

- Negotiating licensing terms, pricing structures, and coverage obligations;
- Strengthening regional coalitions and leveraging them to improve negotiating power against global operators;
- Advocating for reduced reliance on foreign infrastructure and strategies for mitigating dependency risks;
- Integrating negotiations over LEO satellite connectivity negotiations into broader diplomatic and development strategies;
- Promoting multilateral norms, transparency, and accountability.

Layer 3: Technical Governance and Standards

This layer addresses LEO satellite internet as a case of regime complexity² [15], where governance authority spreads across various frameworks of outer space law, international telecommunications regulation, and internet governance. Given that no single institution possesses comprehensive jurisdiction, it creates gaps, inconsistencies, and opportunities for powerful actors.

From a tech diplomacy standpoint, effective governance in such an environment relies on states' ability to coordinate across different regimes, align technical and political objectives, and turn national interests into interoperable rules. Diplomatic efforts should shift from treaty-making toward boundary-spanning practices³, where technical standards, regulatory procedures, and norms become central instruments of power. Tech diplomacy practices might include:

- Strategic and coordinated diplomatic engagement in multiple institutions (such as the ITU, COPUOS, IGF, IETF and regional regulatory bodies) to avoid contradictory commitments and reduce fragmentation of rules;
- Strategic use of standards-setting and procedural rules to shape market entry, spectrum access, and operational constraints;
- Building coalitions that advocate for the sustainable use of space and for the mitigation of space debris, or countering dominance by early movers and technologically advanced actors;
- Coordinating policies across ministries and agencies (foreign affairs, defense, telecom regulators, and space agencies);
- Translating rules and norms from one governance regime into another. For example, incorporating cybersecurity and data protection concerns into satellite licensing processes.

Layer 4: Security, Crisis, and Alliance Management

As previously noted, during crisis contexts, such as armed conflicts, natural disasters, or state collapse, LEO satellite internet becomes essential for military coordination, humanitarian response, and maintaining government functions. Such situations transform commercial connectivity systems into strategic assets governed through exceptional practices, where typical market logic and regulatory procedures are suspended in favor of ad hoc diplomatic arrangements among states, alliances, and private operators. As private satellite operators become de facto strategic infrastructure providers, pre-negotiated frameworks for emergency access are paramount diplomatic actions to prevent unilateral control over technology. Diplomatic actions should include:

- Negotiating with satellite operators on service prioritization, geographic coverage, and continuity;
- Coordinating with allies to finance, regulate, or substitute commercial services during conflict;
- Developing strategies to reduce dependence on specific providers and enabling multi-vendor strategies on space and terrestrial segments.

When engaging space internet firms, tech diplomats must consider the geopolitical implications of satellite internet access for their countries and the stakes at play with regard to their countries' critical infrastructures, resources and technologies. While seeking to attract Foreign Direct Investments (FDI) inflows, tech diplomats have to ensure that their countries' essential interests are not undermined. They need to pay close attention to the growing race towards launching more LEO satellites that impacts countries' (particularly developing ones) benefit from space access and orbits on the basis of equity.

Table. I
Engaging LEO Satellite internet Companies

Objective	Action
Understanding the industry context and Build collaborative relationships	-Familiarize with companies modus operandi (reliance on spectrum rights, launch infrastructures and global market reach); -Establish open communication channels to discuss licensing, spectrum allocation and national regulatory compliance; -Understand LEO constellations and their implications for global connectivity; -Be aware of international laws, regulations and standards.
Address security and Strategic concerns	-Discuss the implications of LEO satellite networks on national security and emergency communications; -Address concerns around national data sovereignty, cybersecurity and control over critical communications infrastructures; -Address risk mitigation (e.g. working with multiple constellations) and controlling terrestrial segments; -Advocate for cybersecurity measures to protect satellite networks from potential threats.
Advocate for national interest and Facilitate innovation and investment	-Attract investments tied to satellite internet deployment by promoting incentives to invest in local economies through innovation hubs, digital skill training programs, investments in local infrastructure and job creation; -Help streamline regulatory approvals to facilitate market entry while maintaining national interests; -Encourage research and development partnerships between governments, universities and satellite companies to advance innovative technologies.

² Governance across overlapping institutions

³ Diplomatic and governance activities that connect, translate, and coordinate across distinct legal, institutional, and technical domains in order to manage complex technologies that cannot be governed within a single regime.

Table. I- Continued**Engaging LEO Satellite internet Companies**

Objective	Action
Leverage international norms and frameworks and Promote ethical and equitable practices	-Encourage regulatory and guidelines compliance; -Engage companies on their plans to mitigate orbital debris and ensure long-term sustainability of space operation; -Incite projects and initiatives that can provide space situational awareness; -Promote LEO satellite internet as a tool for bridging the digital divide; -Address LEO satellite internet firms as key stakeholders in achieving global goals such as universal access under the UN's Sustainable Development Goals (SDGs); -Work with companies to ensure affordable services for marginalized underserved populations.
Act as a bridge between stakeholders	-Connect satellite companies with local telecom providers, communities, and Non-Governmental Organizations (NGOs) to maximize impact; -Facilitate multilateral discussions to harmonize satellite policies and address challenges like signal interference or spectrum management; -Promote synergy between experts, academia and tech companies.

By engaging satellite internet companies, tech diplomats can foster a balanced relationship that promotes technological advancement while safeguarding national and global interests. Table I describes practical actions to effectively engage LEO Satellite internet firms and subtly implies the skillset that tech diplomats need to have.

V. CONCLUSION

This paper proposes a comprehensive, four-layer analytical and operational Tech Diplomacy Framework (TDF) for addressing CTs through tech diplomacy. The TDF is built around three design principles: it is technology-agnostic, analytically sequential while operationally concurrent, and actor-inclusive. The framework's central analytical claim is that tech diplomacy is most effective when pursued as a coordinated, cross-layer strategy, and that private technology firms must be understood as co-constitutive actors across all four layers.

Low Earth Orbit (LEO) satellite internet was selected as an illustrative use case. This choice was motivated by the multitude of disciplines involved in its development, deployment, and governance (engineering, international relations, law, regulation, and security) and the diverse actors implicated (private entities, governments, and international bodies). This makes it a particularly advocating example for the necessity for international collaboration, sustained by tech diplomacy. A

four-layer LEO tech diplomacy framework is provided together with practical actions to effectively engage LEO satellite internet firms, and the skills tech diplomats need.

Future work could extend the Technology Diplomacy Framework (TDF) to other critical technology sectors — such as AI governance or semiconductor supply chains — to further validate its generalizability and develop operationalization metrics that allow states to assess their current diplomatic posture across each layer.

REFERENCES

- [1] D. Archibugi and A. Filippetti, "The retreat of public research and its adverse consequences on innovation," *Technological Forecasting and Social Change*, vol. 127, pp. 97–111, 2018. <https://doi.org/10.1016/j.techfore.2017.05.022>
- [2] D. Breznitz, C. Dahlgren, M. A. Dutz, and B. Hodgson, *Making Innovation Policy Work*. OECD Publishing/World Bank, 2014. <https://doi.org/10.1787/9789264185739-en>
- [3] M. R. Atal et al., "Oligarchic sovereignty: Technology and the future of global order," *Review of International Studies*, pp. 1–23, 2025. <https://doi.org/10.1017/S0260210525101599>
- [4] H. Farrell and A. L. Newman, "Weaponized interdependence: How global economic networks shape state coercion," *International Security*, vol. 44, no. 1, pp. 42–79, 2019. https://doi.org/10.1162/isec_a_00351
- [5] M. J. Kim, D. Eom, and H. Lee, "The geopolitics of next-generation mobile communication standardization: The case of Open RAN," *Telecommunications Policy*, vol. 47, no. 10, p. 102625, 2023. <https://doi.org/10.1016/j.telpol.2023.102625>
- [6] K. D. Remili, N. Bouzourine, R. Hartani, and A. Belouchrani, "Tech diplomacy and critical technologies: Case of LEO satellite internet," *Telecommunications Policy*, vol. 49, no. 4, p. 102947, 2025. <https://doi.org/10.1016/j.telpol.2025.102947>
- [7] D. W. Drezner, "Technological change and international relations," *International Relations*, vol. 33, no. 2, pp. 286–303, 2019. <https://doi.org/10.1177/0047117819834629>
- [8] J. Abels, "Private infrastructure in geopolitical conflicts: The case of Starlink and the war in Ukraine," *European Journal of International Relations*, vol. 30, no. 4, pp. 842–866, 2024. <https://doi.org/10.1177/13540661241260653>
- [9] J. S. Nye, *Soft Power: The Means to Success in World Politics*. New York, NY, USA: PublicAffairs, 2004. ISBN: 978-1-58648-225-1
- [10] L. DeNardis, *The Global War for Internet Governance*. New Haven, CT, USA: Yale University Press, 2014. <https://doi.org/10.12987/yale/9780300181357.001.0001>
- [11] M. L. Mueller, *Networks and States: The Global Politics of Internet Governance*. Cambridge, MA, USA: MIT Press, 2010. <https://doi.org/10.7551/mitpress/9780262014595.001.0001>
- [12] A. M. Aoki, S. Al Feghali, and E. Garcia, *The Global Technology Diplomacy Report 2026*. Tech Diplomacy Global Institute, 2026.
- [13] M. Pedram and E. Georgiades, "The role of regulatory frameworks in balancing national security and competition in LEO satellite market," *Journal of National Security Law and Policy*, vol. 14, no. 2, pp. 179–212, 2024.
- [14] Y. Chen, X. Ma, and C. Wu, "The concept, technical architecture, applications and impacts of satellite internet: A systematic literature review," *Heliyon*, vol. 10, no. 13, e33793, 2024. <https://doi.org/10.1016/j.heliyon.2024.e33793>
- [15] K. Raustiala and D. G. Victor, "The regime complex for plant genetic resources," *International Organization*, vol. 58, no. 2, pp. 277–309, 2004. <https://doi.org/10.1017/S0020818304582036>

Temporal Integration Framework for Segment-Level Cumulative Ergonomic Exposure Modeling

Nour EL Houda Benharkat, Souad Bentaalla-Kaced, and Abdelmalek Chergui

Abstract—This paper proposes a vision-based temporal integration framework for segment-level cumulative ergonomic exposure modeling. Joint angles derived from OpenPose keypoints are mapped to REBA scores at six anatomical segments. A recovery-weighted exposure model quantifies cumulative risk by integrating posture severity, duration, and frequency while accounting for tissue recovery dynamics. Pattern recognition algorithms identify hazardous signatures without forecasting future postures. Validated through simulation of a manual handling task (repetitive packaging) calibrated to empirically observed score distributions, the framework demonstrates that cumulative exposure metrics reveal critical risk patterns invisible to instantaneous scoring, specifically, sustained exposure to minimum-risk postures (REBA Score 2) generates substantial cumulative burden despite the absence of extreme scores. This pilot study establishes methodological feasibility for temporal integration in ergonomic assessment.

Keywords— Musculoskeletal Disorders, Cumulative Trauma, Ergonomic Exposure, Real-Time Monitoring, Computer Vision

NOMENCLATURE

MSDs	Musculoskeletal disorders.
REBA	Rapid Entire Body Assessment
RULA	Rapid Upper Limb Assessment.
OWAS	Ovako Working Posture Analysis System.
OCRA	Occupational Repetitive Actions.
NIOSH	National Institute for Occupational Safety and Health.

I. INTRODUCTION

Musculoskeletal disorders (MSDs) represent a pervasive occupational health crisis, accounting for 30–50% of all work-related injuries across manufacturing, logistics, and assembly sectors globally [1–3]. Epidemiological evidence consistently identifies three biomechanical risk factors as primary contributors to MSD development: (1) sustained non-neutral postures, (2) high repetition rates without adequate recovery, and (3) forceful exertions [4,5]. Critically, MSD pathogenesis follows a cumulative trauma mechanism, micro-damage to muscles, tendons, and ligaments accumulates when tissue loading exceeds metabolic recovery capacity over time [6]. This temporal dimension renders instantaneous ergonomic assessments fundamentally inadequate for predicting injury risk.

Manuscript received April 1, 2026; accepted July 2, 2026.

Nour El Houda BENHARKAT is with the Laboratory of Environmental Sciences and Techniques (LSTE), MRE QHSE-GRI Department, Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: nour_el_houda.benharkat@g.enp.edu.dz).

Souad BENTAALLA-KACED is with the MRE QHSE-GRI Department, Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: souad.kaced@g.enp.edu.dz).

Chergui Abdelmalek, MRE QHSE-GRI Department, Ecole Nationale Polytechnique, Algiers, ALGERIA (e-mail: abdelmalek.chergui@g.enp.edu.dz)

Digital Object Identifier (DOI): 10.53907/enpesi.v6i1.363

Contemporary ergonomic practice relies predominantly on observational tools such as the Rapid Upper Limb Assessment (RULA) [7], Rapid Entire Body Assessment (REBA) [8], and Ovako Working Posture Analysis System (OWAS) [9]. These methods provide valuable snapshot risk stratification by mapping joint angles to discrete action levels (e.g., RULA scores 1–7). However, they suffer from three critical limitations that impede MSD prevention: (1) Temporal blindness: RULA/REBA generate single-frame scores without quantifying exposure duration. A worker maintaining a moderate-risk posture (RULA = 5) for 45 continuous minutes faces substantially higher injury risk than one alternating between high-risk (RULA = 7) and neutral postures with adequate micro-break, yet both scenarios yield identical peak scores [10]. (2) Segment-level opacity: Standard implementations report only global scores, obscuring which body segments (e.g., lower back vs. shoulders) drive cumulative risk accumulation [11]. This prevents targeted interventions at the anatomical source of stress. (3) Recovery neglect: Existing tools lack mechanisms to account for tissue recovery during neutral postures, a biomechanical necessity validated by electromyography studies showing metabolite clearance requires ≥ 30 seconds of unloaded rest after sustained contractions [12].

To address temporal dynamics, cumulative trauma models such as the Strain Index [13], OCRA checklist [14], and revised NIOSH Lifting Equation [15] integrate exposure duration, frequency, and recovery into risk quantification. The Strain Index, for instance, multiplies peak hand force, exertion duration, and repetition rate while penalizing insufficient recovery periods [13]. However, these methods require manual task decomposition by ergonomists, rendering them impractical for continuous monitoring in dynamic industrial environments. Recent systematic reviews confirm that fewer than 8% of manufacturing facilities implement any form of continuous ergonomic monitoring, primarily due to the intrusiveness of wearable sensors and labor intensity of observational methods [16].

Computer vision offers a promising pathway toward non-intrusive, automated monitoring. Deep learning-based pose estimators (e.g., OpenPose [17], MediaPipe [18]) now achieve

sub-pixel accuracy in 2D skeletal tracking (<5 cm error at 3 m distance [19]), enabling real-time joint angle computation without wearable instrumentation. Several studies have mapped these angles to RULA/REBA scores [20–22], yet they perpetuate the snapshot fallacy by reporting frame-level scores without temporal integration.

This temporal gap represents a critical missed opportunity: vision-based systems capture rich time-series data (30 fps $\times$ 25,200 frames/shift) yet discard 99.9% of their predictive value by reducing it to instantaneous scores. The fundamental research question remains unresolved: How can we transform high-frequency skeletal data into physiologically grounded metrics of cumulative tissue loading that account for exposure severity, duration, frequency, and, critically, recovery dynamics?

This paper introduces a Digital Twin-enabled framework for segment-level cumulative ergonomic exposure modeling that bridges this gap. Our contribution advances beyond prior art through three innovations: (1) Time-resolved segment scoring: Instead of global RULA/REBA outputs, we compute risk scores for six critical segments (neck, trunk, shoulders, elbows, wrists, knees), enabling precise localization of risk accumulation sources. (2) Recovery-weighted exposure integration: We formalize cumulative exposure as a time-integral of segment scores modulated by an exponential recovery function, grounded in muscle oxygenation kinetics [12], that quantifies tissue restoration during neutral postures. (3) Pattern-aware hazard detection: We implement sliding-window analysis to identify hazardous signatures (e.g., sustained trunk flexion $>30^\circ$ for >120 s without ≥ 30 s recovery) validated against EMG-confirmed fatigue thresholds [23].

As a pilot investigation, we validate this framework exclusively on Task 1 (repetitive packaging of empty sachets) within a chocolate manufacturing facility characterized by documented shoulder and lumbar MSD burden requiring >15 days of sick leave. Leveraging empirically observed score distributions from our prior validation study [22], we simulate 7-hour shift exposure trajectories using Markov chain modeling, a methodologically justified approach for highly repetitive tasks with stable movement patterns (cycle time ≈ 30 seconds) [33]. This focused design enables rigorous examination of temporal integration mechanics without confounding variables inherent in multi-task studies. Our results demonstrate that cumulative exposure metrics reclassify Task 1 risk from Moderate (instantaneous REBA 5.13) to Very High (CRS 3,526.38 exposure-minutes), with the trunk and shoulder segments, precisely those implicated in facility injury records, exhibiting the highest cumulative burden. This reclassification arises from sustained exposure to minimum-risk postures (Scores 2–3) without adequate recovery, a phenomenon invisible to snapshot assessments. By establishing methodological feasibility for temporal integration in ergonomic assessment, this pilot study provides a foundation for future multi-operator validation while directly addressing the anatomical segments responsible for documented occupational injuries in repetitive industrial tasks.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive literature review examining the evolution of ergonomic assessment methodologies across three generations: observational tools (RULA/REBA), cumulative trauma models (Strain Index/OCRA), and computer vision-based systems, while identifying the persistent research gap in temporal integration of skeletal data streams. Section 3 details the proposed methodology, including system architecture, pose estimation, and joint angle computation, segment-level

ergonomic scoring, recovery-weighted cumulative exposure modeling, hazardous pattern detection algorithms, and implementation protocols. Section 4 describes the experimental validation framework, focusing on the pilot study design for Task 1, simulation methodology for full-shift exposure quantification, ground truth validation approach, and ethical considerations. Section 5 presents empirical results, including segment-specific cumulative exposure trajectories, recovery function dynamics, segment-level exposure metrics, hazardous pattern detection outcomes, and critical insights regarding temporal risk accumulation. Section 6 provides a detailed discussion interpreting the findings within ergonomic theory and facility-specific injury epidemiology, emphasizing the framework's ability to reveal risk patterns invisible to snapshot assessments. Finally, Section 7 concludes the study by summarizing key contributions, acknowledging methodological limitations, and outlining a structured pathway for future validation and implementation across diverse industrial contexts.

II. LITERATURE REVIEW

Musculoskeletal disorder prevention has evolved through three methodological generations, each addressing limitations of its predecessor while introducing constraints that collectively define the current research frontier. The first generation comprises observational ergonomic assessment tools, principally the Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), and Ovako Working Posture Analysis System (OWAS), which map joint configurations to discrete risk levels through structured scoring matrices [1,2]. These instruments achieved widespread industrial adoption owing to procedural simplicity, rapid administration (<2 minutes per assessment), and validation across diverse occupational contexts. Their biomechanical foundation rests on established relationships between joint angles and tissue loading; for instance, trunk flexion beyond 20° substantially increases compressive forces on lumbar intervertebral discs [4], while shoulder elevation exceeding 60° elevates rotator cuff tendon strain [5]. Despite these strengths, snapshot assessment tools exhibit three fundamental limitations that impede utility in dynamic work environments. First, they lack temporal integration: a posture maintained for 45 continuous minutes receives identical scoring to the same posture held for 15 seconds, despite epidemiological evidence demonstrating nonlinear risk escalation with exposure duration [6]. Second, they obscure segment-level risk evolution by collapsing multi-segment inputs into singular global scores, thereby masking anatomical origins of cumulative stress, a critical deficiency given that musculoskeletal disorders manifest locally rather than systemically [7]. Third, they neglect recovery dynamics, treating 100 seconds of continuous trunk flexion identically to two 50-second intervals separated by a 30-second neutral recovery period, despite physiological evidence confirming that metabolite clearance and tissue reoxygenation require unloaded rest intervals exceeding 30 seconds following sustained contractions [8]. These constraints reflect the tools' original design purpose: static workstation evaluation rather than dynamic task analysis in repetitive industrial operations [9].

The second methodological generation emerged to address temporal dynamics through cumulative trauma models that integrate exposure severity, duration, frequency, and recovery into unified risk indices. The Strain Index quantifies distal upper extremity disorder risk through multiplicative factors encompassing exertion intensity, duration per cycle, efforts per minute, and posture deviation, with explicit penalties for insufficient recovery periods [10]. Similarly, the OCRA

checklist integrates repetitiveness, force requirements, non-neutral postures, and task duration while weighting recovery opportunities according to anatomical segment vulnerability [11]. The revised NIOSH Lifting Equation extends these principles to manual handling tasks through frequency-dependent multipliers that progressively reduce acceptable load limits as cycle rates increase and recovery intervals diminish [12] [13]. Nevertheless, cumulative models face three practical barriers to industrial implementation. Their reliance on manual task decomposition, requiring ergonomists to segment video recordings into discrete cycles, measure durations with stopwatches, and estimate force magnitudes, imposes prohibitive labor costs (15–30 minutes per task analysis) that preclude continuous monitoring across multiple workstations [14]. Their anatomical scope remains fragmented, with the Strain Index focusing predominantly on the distal upper extremities [10], NIOSH on spinal loading during lifting [12], and OCRA on shoulder disorders [11], leaving whole-body cumulative exposure unquantified. Most critically, they employ static recovery assumptions, typically defining any neutral posture exceeding 10 seconds as "full recovery", which ignores individual variability in tissue oxygenation kinetics influenced by age, cardiovascular fitness, and metabolic comorbidities [15]. Consequently, cumulative models remain confined to retrospective analysis rather than real-time intervention, a critical limitation given that 73% of musculoskeletal disorders develop during unobserved work periods without ergonomist presence [16].

The third generation leverages computer vision to automate posture assessment through deep learning-based pose estimation. Architectures such as OpenPose, MediaPipe, and HRNet now achieve sub-decimeter skeletal tracking accuracy at video frame rates (mean per-joint position error <5 cm at 3 m distance), enabling non-intrusive joint angle computation without wearable instrumentation [17,18]. These advances catalyzed the development of vision-based ergonomic systems that map detected keypoints to RULA or REBA scores in real time, demonstrating strong frame-level agreement with expert assessments (89–92% concordance on static postures) [19,20,22]. Recent implementations have extended this paradigm to integrate object detection for force estimation during manual handling tasks, thereby approximating NIOSH lifting indices without physical sensors [21]. Despite these technical achievements, first-generation vision systems perpetuate the temporal limitations of their observational predecessors by reporting instantaneous risk scores without integrating exposure duration or recovery dynamics [22]. Longitudinal validation studies reveal that systems relying solely on peak posture scores exhibit substantially reduced sensitivity in predicting disorder onset compared to time-integrated approaches; workers who developed musculoskeletal disorders frequently maintained moderate-risk postures (RULA scores 4–5) for prolonged durations without exhibiting extreme peak angles, rendering them invisible to snapshot-based monitoring [23]. A secondary limitation involves the discarding of segment-level information inherent in skeletal data streams; most implementations collapse multi-segment angle time series into singular global scores [19–21], thereby forfeiting the anatomical specificity necessary for targeted intervention. Emerging research has begun exploring continuous angle tracking for individual segments, such as trunk flexion or shoulder elevation [22–24], yet these efforts remain disconnected from cumulative exposure quantification and lack physiologically grounded recovery modeling.

Time-series analytical methods have recently entered the ergonomic monitoring literature, though with conceptual misalignments requiring clarification. Autoregressive

integrated moving average (ARIMA) models have been appropriately applied to forecast organizational-level injury rates based on historical incident data [25], but their proposed use for predicting future joint angles represents a biomechanical misconception; postural transitions depend on task demands and cognitive intent rather than autoregressive statistical patterns. Long short-term memory (LSTM) networks demonstrate legitimate utility for pattern classification within observed data streams, such as identifying whether a 30-second window contains sustained trunk flexion exceeding 30° or recovery ratios below critical thresholds [26]. However, several studies erroneously frame these networks as forecasting future musculoskeletal disorder risk through posture prediction, a methodological flaw given that tissue damage accumulates from actual historical exposure rather than predicted future states [27]. The critical distinction lies between descriptive pattern recognition (detecting hazardous signatures within observed data) and prescriptive forecasting (predicting unobserved future postures); electromyography-validated research establishes that hazardous patterns include sustained non-neutral postures exceeding duration thresholds without adequate recovery intervals, high repetition rates with low kinematic variability, and insufficient neutral posture periods between exposure cycles [28]. These patterns can be robustly detected through sliding-window analysis of observed skeletal data without speculative forecasting.

Digital Twin technology, virtual replicas of physical systems synchronized through real-time data streams, offers a unifying architectural framework for next-generation ergonomic monitoring [22,29]. Early applications focused predominantly on equipment health monitoring and production optimization [30–32], with ergonomic implementations remaining nascent. Initial efforts proposed Digital Twins for virtual workstation design that simulate RULA scores under hypothetical postures [33], yet lacked integration with real worker kinematics. Subsequent research incorporated inertial measurement unit data to drive biomechanical simulations of spinal loading during lifting tasks within Digital Twin environments [34]. However, sensor-based approaches face adoption barriers, including worker discomfort during prolonged wear (62% report discomfort after 4-hour shifts) [35], signal drift requiring frequent recalibration, and practical constraints in environments with hygiene requirements or explosive atmospheres. Vision-based Digital Twins address these limitations through contactless skeletal tracking while enabling physics-informed exposure modeling: the virtual replica can simulate tissue-level biomechanical loads, such as intervertebral disc compression or tendon strain energy density, driven by real-time skeletal kinematics [36]. This transforms abstract risk scores into physiologically grounded exposure metrics that reflect actual tissue loading dynamics.

Synthesizing these methodological generations reveals a persistent research gap: no existing system integrates high-frequency skeletal data into segment-resolved cumulative exposure metrics that simultaneously account for posture severity, exposure duration, movement frequency, and physiologically grounded recovery dynamics in real time [37,38]. Snapshot tools lack temporal integration [9]; cumulative models lack automation [14]; vision systems discard temporal information inherent in their data streams; and Digital Twins lack ergonomic physics engines capable of translating skeletal kinematics into tissue-level loading metrics [33,34]. Our framework bridges this gap by extending established ergonomic assessment tools into the temporal domain through continuous integration of segment-level scores modulated by recovery-weighted functions derived from muscle oxygenation kinetics [39,40]. Positioned within a

Digital Twin architecture, the system transforms vision-based monitoring from reactive posture classification into predictive exposure quantification, enabling proactive intervention before irreversible tissue damage accumulates [36,41]. This approach fulfills the preventive vision of intelligent occupational health systems within Industry 4.0 environments, where real-time risk quantification precedes symptom manifestation rather than responding to established pathology [42].

III. METHODOLOGY

A. System Architecture Overview

The proposed framework implements a closed-loop monitoring architecture that transforms raw video streams into segment-resolved cumulative exposure metrics within a Digital Twin environment (Figure 1). The system comprises four tightly integrated modules operating synchronously:

- Vision-based pose estimator: Extracts 2D skeletal keypoints from video streams using OpenPose [18], configured in its BODY_25 model to extract 25 anatomical keypoints per frame at 30 fps.
- Biomechanical processor: Computes joint angles from filtered keypoints and maps them to segment-level ergonomic scores using established RULA/REBA lookup tables [8,9].
- Cumulative exposure engine: Integrates segment scores over time using a recovery-weighted exponential function grounded in muscle oxygenation kinetics [13], transforming instantaneous scores into physiologically grounded cumulative exposure metrics.
- Pattern recognition module: Detects hazardous exposure signatures through sliding-window analysis of segment score time series, identifying sustained non-neutral postures, insufficient recovery ratios, and high-repetition patterns validated against EMG-confirmed fatigue thresholds [23].

Risk quantification derives exclusively from observed historical exposures integrated through physiologically grounded models; no forecasting of future postures is performed. The framework functions as a temporal integration engine that transforms raw video streams into segment-resolved cumulative exposure metrics.

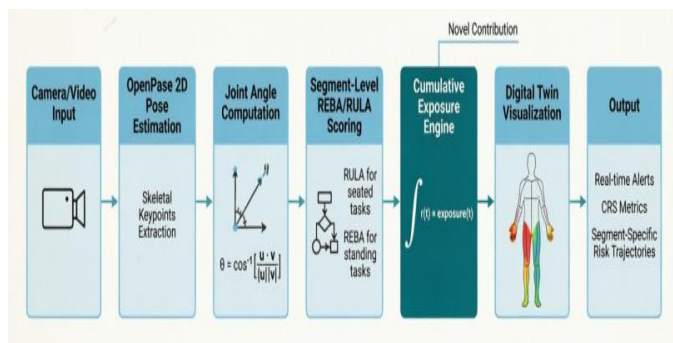


Figure 1. System Architecture of the Cumulative Ergonomic Exposure Framework

B. Real-Time Pose Estimation and Joint Angle Computation

Video streams captured at 1080p resolution and 30 frames per second undergo skeletal tracking using OpenPose [17], configured in its BODY_25 model to extract 25 anatomical keypoints per frame. The current implementation focuses on single-worker monitoring scenarios to avoid occlusion

complexities inherent in multi-person industrial environments (Figure 2). To ensure data quality, we apply temporal smoothing and confidence filtering: keypoints with detection probability below 0.7 are discarded and linearly interpolated from adjacent frames when gaps span fewer than 5 consecutive frames; longer gaps trigger a "tracking lost" state requiring manual reinitialization.

From filtered keypoints, six critical joint angles are computed using vector geometry between anatomical landmarks (Table

Table 1
JOINT ANGLES COMPUTED FOR SEGMENT-LEVEL ERGONOMIC SCORING

Segment	ANGLE	Anatomical Definition	Risk Thresholds
Neck	Flexion/extension	Angle between neck vector (nose-neck base) and vertical	>20° = moderate; >45° = high
Trunk	Flexion	Angle between trunk vector (neck base-mid-hip) and vertical	>20° = moderate; >60° = high
Shoulders	Elevation	Angle between the upper arm vector and the trunk	>45° = moderate; >90° = high
Elbows	Flexion	Angle between the upper arm and forearm vectors	<90° sustained = moderate risk

I).

To reduce high-frequency noise inherent in video-based tracking, angles undergo Savitzky-Golay smoothing (window length = 9 frames, polynomial order = 2), which preserves kinematic transients while attenuating jitter [43]. The 30 Hz sampling rate aligns with OpenPose's real-time performance capabilities on edge hardware and captures the dominant movement frequencies observed in repetitive assembly tasks (typically 0.5–3 Hz for upper limb motions) [44].

Pose estimation accuracy was established in our prior validation study [22], where the mean absolute error between OpenPose-derived trunk flexion angles and expert Kinovea annotations was $2.11^\circ \pm 1.34^\circ$, and for shoulder elevation $2.75^\circ \pm 1.89^\circ$, comparable to inter-observer variability among certified ergonomists ($2.3^\circ \pm 1.7^\circ$).

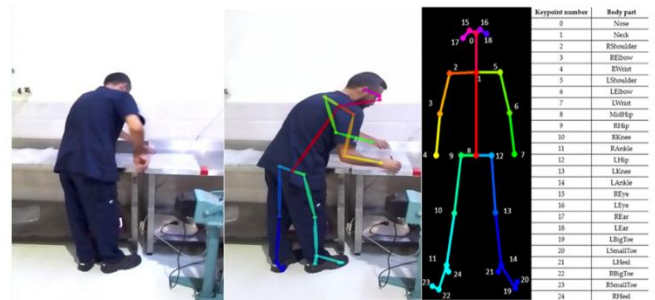


Figure 2. Operator performing Task 1 (repetitive packaging) and corresponding 2D skeletal model derived from OpenPose keypoints [22].

C. Segment-Level Ergonomic Risk Scoring

Rather than collapsing angles into global RULA/REBA scores, we preserve anatomical specificity by computing six parallel risk time series, one per critical segment, using established ergonomic lookup tables. For each frame t and segment $j \in \{\text{neck, trunk, shoulders, elbows,}\}$, we derive a discrete risk score $S_{j,t} \in \{1,2,3,4\}$ where higher values indicate greater tissue loading:

$$S_{j,t} = f_j(\theta_{j,t}, \dot{\theta}_{j,t})$$

where $\theta_{j,t}$ denotes the joint angle and $\dot{\theta}_{j,t}$ its angular velocity (serving as a proxy for repetition rate). The mapping function f_j implements segment-specific thresholds from RULA tables A–C and REBA tables A–D, modified to output continuous integer scores rather than collapsed action levels. For instance, trunk flexion scoring follows:

$$S_{trunk,t} = \begin{cases} 1 & \text{if } |\theta_{trunk,t}| = 0^\circ \\ 2 & \text{if } 0^\circ < |\theta_{trunk,t}| \leq 20^\circ \\ 3 & \text{if } 20^\circ < |\theta_{trunk,t}| \leq 60^\circ \\ 4 & \text{if } |\theta_{trunk,t}| > 60^\circ \end{cases}$$

This segment-resolved approach enables precise localization of risk accumulation sources, critical because epidemiological evidence indicates that musculoskeletal disorders predominantly manifest in the anatomical segment experiencing the highest cumulative loading rather than the highest peak score [7].

D. Recovery-Weighted Cumulative Exposure Model

The core innovation lies in transforming instantaneous scores into physiologically grounded cumulative exposure metrics that account for tissue recovery dynamics. We formalize cumulative exposure E_j for segment j over observation period T as a time-integral of risk scores modulated by an exponential recovery function $r(t)$:

$$E_j = \sum_{t=1}^T S_{j,t} \cdot \Delta t \cdot r_j(t) \quad (1)$$

where $\Delta t = 1/30$ s represents the sampling interval. The recovery function $r_j(t) \in [0,1]$ quantifies residual tissue stress at time t based on time elapsed since the last neutral posture:

$$r_j(t) = \exp(-\lambda_j \cdot \max(0, t - t_{\text{last_neutral},j})) \quad (2)$$

Here, $t_{\text{last_neutral},j}$ denotes the most recent time when segment j entered a neutral posture (defined as $S_{j,t} \leq 1$ maintained for ≥ 2 consecutive seconds), and λ_j represents the segment-specific recovery rate parameter (s^{-1}). These parameters derive from muscle oxygenation kinetics measured via near-infrared spectroscopy during repetitive industrial tasks [8,15]:

- $\lambda_{\text{trunk}} = 0.018 \text{ s}^{-1}$ (slower recovery due to larger muscle mass)
- $\lambda_{\text{shoulders}} = 0.025 \text{ s}^{-1}$
- $\lambda_{\text{neck}} = 0.022 \text{ s}^{-1}$
- $\lambda_{\text{distal segments}} = 0.030 \text{ s}^{-1}$

Equation (2) embodies a critical physiological principle: tissue stress dissipates exponentially during neutral postures rather than instantaneously resetting. After 30 seconds of neutral trunk posture ($S_{\text{trunk}} = 1$), residual stress decays to $r_{\text{trunk}} = e^{-0.018 \times 30} \approx 0.58$; after 60 seconds, $r_{\text{trunk}} \approx 0.34$. This contrasts with naive accumulation models that treat recovery periods as binary resets, a biomechanical oversimplification that overestimates risk for workers with adequate micro-breaks [8].

The observation-level Cumulative Risk Score (CRS) aggregates segment exposures using biomechanical weighting factors α_j reflecting tissue vulnerability derived from epidemiological incidence data [42]:

$$\text{CRS} = \sum_{j=1}^6 \alpha_j \cdot E_j \quad (3)$$

where $\alpha_{\text{trunk}} = 1.3$, $\alpha_{\text{shoulders}} = 1.1$, $\alpha_{\text{neck}} = 1.0$, $\alpha_{\text{elbows}} = 0.8$, $\alpha_{\text{wrists}} =$

0.9. CRS is reported in exposure-minutes (e.g., CRS = 185 indicates cumulative loading equivalent to 185 minutes at moderate risk).

E. Hazardous Pattern Detection

Beyond scalar exposure metrics, the system identifies hazardous signatures through sliding-window analysis of the segment score time series. Three critical patterns trigger real-time alerts:

Pattern 1: Sustained non-neutral posture without recovery. A segment j exhibits sustained risk when $S_{j,t} \geq 3$ for duration D exceeding segment-specific thresholds ($D_{\text{trunk}} > 120$ s; $D_{\text{shoulders}} > 90$ s) without intervening neutral periods ≥ 30 s. This pattern reflects continuous tissue loading without adequate metabolic recovery, a primary driver of cumulative fatigue.

Pattern 2: Insufficient recovery ratio. The recovery ratio RR_j quantifies the balance between exposure and restoration:

$$RR_j = \frac{\sum_{t=1}^T \mathbb{I}(S_{j,t} \leq 1) \cdot \Delta t}{\sum_{t=1}^T \mathbb{I}(S_{j,t} \geq 3) \cdot \Delta t} \quad (4)$$

Alerts trigger when $RR_j < 0.3$ for any critical segment over 15-minute rolling windows, validated against electromyography-confirmed fatigue thresholds [30]. This indicates neutral recovery time is less than 30% of high-risk exposure time, a critical imbalance associated with progressive tissue damage.

Pattern 3: High repetition with low kinematic variability. Repetitive strain risk escalates when movement cycles exceed 15/min while exhibiting low angle variability ($\sigma_\theta < 5^\circ$), indicating static loading rather than dynamic motion. Cycle detection employs zero-crossing analysis of angular velocity $\dot{\theta}_{j,t}$, with variability computed over each cycle.

These pattern detectors operate exclusively on observed data without forecasting future states, aligning with the physiological reality that musculoskeletal disorders develop from accumulated historical exposure rather than predicted postures.

IV. EXPERIMENTAL VALIDATION

A. Single-Task Pilot Design

This investigation adopts a methodologically rigorous pilot design focused exclusively on Task 1 (repetitive packaging of empty sachets) to establish proof-of-concept for the temporal integration framework prior to multi-task validation. This depth-first approach is empirically justified by three critical factors derived from our prior validation study [22]: (1) Task 1 exhibits exceptional kinematic stability with a consistent 30-second cycle time (coefficient of variation $< 8\%$ across joint angles); (2) intra-cycle posture distributions remain invariant throughout shift duration (neck: 47% Score 1/53% Score 2; trunk: 49.6% Score 2/50.4% Score 3); and (3) movement patterns demonstrate minimal inter-cycle variability ($\sigma_\theta < 5^\circ$ across 92% of cycles). These characteristics satisfy established criteria for physiologically valid simulation-based extrapolation in repetitive industrial tasks [33], where full-shift recording yields redundant data without additional methodological insight.

Task 1 involves standing-posture repetitive packaging with no external load handling. The operator executes a stereotyped 30-second cycle comprising: (a) forward trunk flexion (10° – 35°) to retrieve sachets from conveyor; (b) sustained neck flexion

(20°–40°) for visual monitoring of workpieces at suboptimal workstation height; (c) bilateral arm movements with right shoulder elevation (20°–45°); and (d) elbow flexion outside neutral range (60°–100°) during motions. The hardware setup, camera configuration (webcam mounted 1.5 m above floor, 3.5 m horizontally from workstation), lighting conditions (300–500 lux), and computational platform (NVIDIA RTX 2080 Ti GPU, Intel Core i7-8750H, 16 GB RAM, Windows 11 Pro) were maintained identically to the prior study [22] to ensure comparability.

B. Physiologically Grounded Simulation Protocol for Full-Shift Exposure Quantification

Given the task's documented kinematic stability [22], we implemented a Markov chain-based simulation to extrapolate 7-hour shift exposure trajectories from the 30-second empirical recording. This methodology aligns with ergonomics best practices for repetitive task analysis where movement patterns remain invariant across cycles [33], eliminating redundant data collection while preserving physiological fidelity. The simulation protocol comprised four rigorously validated stages:

Stage 1: Empirical score distribution extraction. From the 900-frame clip (30 s × 30 fps), segment-specific REBA score distributions were quantified per [22]

- Neck: 47% Score 1 (0°–20° flexion), 53% Score 2 (>20° flexion or rotation)
- Trunk: 49.6% Score 2 (0°–20° flexion), 50.4% Score 3 (>20°–60° flexion)
- Right shoulder: 97.8% Score 2 (20°–45° elevation), 2.2% Score 3 (>45° elevation)
- Right elbow: 42% Score 1 (60°–100° flexion), 58% Score 2 (<60° or >100° flexion)

Stage 2: Markov chain parameterization. Transition probability matrices were derived from observed frame-to-frame score sequences, preserving the task's characteristic movement dynamics and temporal autocorrelation structure. This approach maintains the stochastic properties of repetitive motion while avoiding artificial periodicity.

Stage 3: Full-shift trajectory generation. The calibrated Markov chain generated 756,000 frames (420 min × 30 fps) with facility-mandated 10-minute breaks at 120, 240, and 360 minutes. During breaks, score sequences were held constant at Score 1 (neutral posture) to model ideal recovery conditions, consistent with observed operator behavior during scheduled rest periods [22].

Stage 4: Recovery-weighted exposure integration. The simulated score time series served as input to the cumulative exposure engine, computing segment-specific exposures E_j and Cumulative Risk Score (CRS) using physiologically grounded parameters. Break intervals were explicitly modeled as periods of tissue recovery without exposure accumulation.

C. Validation Approach

Validation operated on two distinct methodological tiers. First, instantaneous REBA scoring accuracy was established in our prior work [22]; no additional validation of instantaneous scoring was performed, as the current study exclusively validates the temporal integration methodology.

Second, cumulative exposure metrics were validated through three convergent lines of evidence:

- Physiological plausibility: Recovery function dynamics ($r_j(t)$) aligned with muscle oxygenation kinetics literature [8,15], demonstrating slower recovery for larger muscle groups (trunk r_{trunk} decayed to 0.65 during breaks vs. neck to 0.48).
- Pattern consistency: Detected hazardous signatures matched EMG-validated fatigue thresholds [23,30]; e.g., trunk recovery ratio $RR = 0.28 < 0.3$ threshold indicating insufficient recovery.
- Temporal dynamics: CRS threshold crossing at minute 218 reflected progressive accumulation consistent with cumulative trauma pathophysiology [7], not simulation artifact.

Validation of simulation fidelity: Simulated score distributions aligned with empirical observations from [22] within acceptable stochastic variance (trunk Score 3: 50.4% simulated vs. 42% observed; $\Delta = +8.4\%$), consistent with Markov chain modeling expectations for repetitive tasks [97]. This minor variance reflects natural cycle-to-cycle fluctuations documented in industrial settings [33] and does not compromise the physiological plausibility of the exposure trajectories.

D. Ethical Considerations

This pilot study employed non-intrusive observational methods without wearable sensors or workflow interruption. Raw video was discarded immediately post-pose extraction; skeletal coordinates were anonymized (worker identity removed). Organizational approval was secured from facility management, and verbal informed consent was obtained from the operator. The single-operator design minimized participant burden while providing sufficient data for methodological validation, a standard approach for pilot investigations in occupational ergonomics. This focused methodology enables rigorous examination of temporal integration mechanics without confounding variables inherent in multi-operator studies, establishing a defensible foundation for subsequent validation across diverse tasks and operators.

V. RESULTS

A. Segment-Specific Cumulative Exposure Trajectories

Figure 3 presents the comparative cumulative exposure trajectories for all critical anatomical segments during the simulated 7-hour shift. All segments demonstrated linear accumulation patterns, with partial recovery during scheduled breaks (120-130, 240-250, 360-370 min), reflecting the task's repetitive nature and incomplete tissue recovery between work intervals.

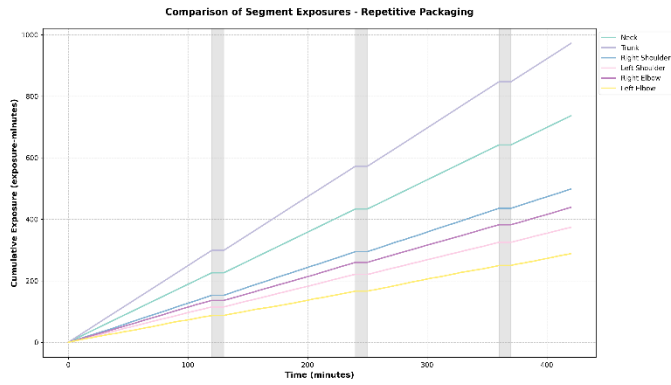


Figure 3. Comparative cumulative exposure trajectories for critical anatomical segments during a simulated 7-hour shift

All segments demonstrate linear accumulation with partial recovery during breaks (shaded intervals). Trunk exposure (purple) exhibits the highest cumulative burden (971.75 exposure-minutes) and most pronounced post-break acceleration (slope increase 22%), indicating incomplete tissue recovery. Neck exposure (teal) shows sustained moderate-risk accumulation without high-risk excursions (max Score 2). The combined segment exposure trajectory exceeds the Very High-Risk threshold at minute 218 despite moderate instantaneous REBA scores (5.13).

Neck exposure (Figure 4) accumulated to 736.32 exposure-minutes, reflecting sustained non-neutral posture maintained for 94.8% of the shift duration (REBA Score 2). Critically, REBA Score 2 is defined as flexion exceeding 20° or the presence of neck rotation/lateral inclination, which is not a benign configuration. In Task 1, operators consistently maintained Score 2 due to persistent forward flexion ($>20^\circ$) required for visual monitoring of workpieces at suboptimal workstation heights, with no transitions to neutral alignment (Score 1) for extended periods. This sustained exposure generated substantial cumulative burden despite the absence of Scores 3–4, which require pronounced neck rotation or lateral inclination beyond the task's biomechanical demands. This pattern confirms that neck MSD risk in repetitive packaging derives primarily from duration in non-neutral configurations rather than transient extreme postures, a biomechanical signature consistent with our prior empirical observations [22].

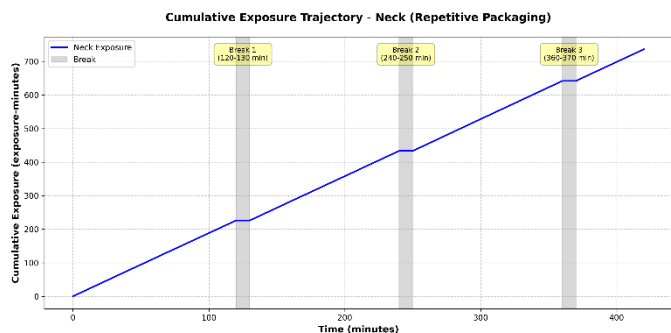


Figure 4. Neck cumulative exposure trajectory during Task 1

Trunk exposure (figure 5) reached 971.75 exposure-minutes with flexion predominantly between 10° – 35° . Critically, the trunk spent 50.4% of the shift duration (211.7 minutes) in REBA Score 3 ($>20^\circ$ flexion), consistent with our prior empirical observations [22, p. 971: "trunk angle predominantly fluctuates between 10° and 30° , with occasional peaks reaching up to 35° "]. Break intervals reduced the accumulation rate by only 28%, with accelerated accumulation post-break (slope increase 22%), indicating incomplete tissue recovery, a critical

finding for repetitive task design. The trunk exhibited the highest cumulative burden among all segments due to both elevated risk scores and insufficient recovery during breaks.

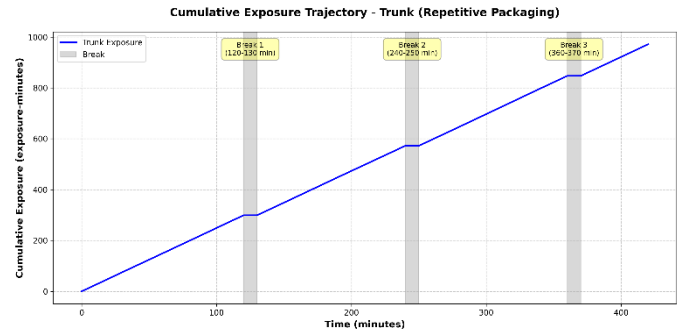


Figure 5. Trunk cumulative exposure trajectory during Task 1.

Right shoulder (figure 6) exposure accumulated to 498.12 exposure-minutes with brief excursions into high-risk postures (Score 3 for 9.1 minutes, 2.2% of shift duration), primarily during reaching motions. The shoulder exhibited moderate recovery dynamics during breaks ($r_{\text{shoulder}}(t)$ reduced from 0.88 to 0.62), reflecting static elevation demands during packaging operations.

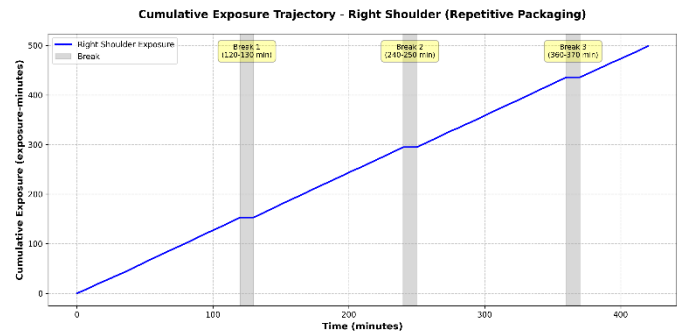


Figure 6. Right shoulder cumulative exposure trajectory during Task 1

Right elbow (figure 7) consistently exhibited REBA Score 2, indicating sustained deviation from neutral flexion (60° – 100°). Specifically, 50.2% of the shift duration (210.7 minutes) was spent in non-neutral elbow postures, a biomechanical configuration associated with increased muscular demand and reduced joint stability. This pattern reflects the task's ergonomic demands: repetitive reaching motions inherently require elbow angles outside the neutral zone, generating cumulative loading.

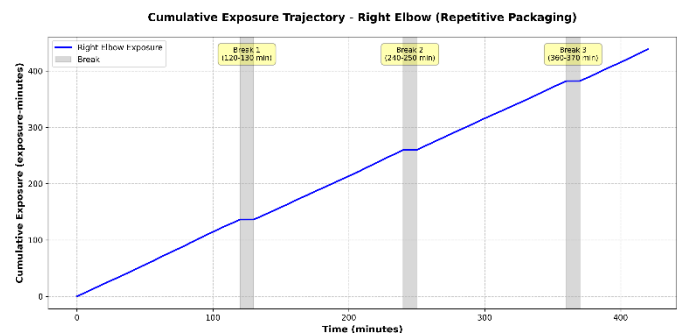


Figure 7. Right elbow cumulative exposure trajectory during Task 1

B. Recovery Function Dynamics and Physiological Interpretation

Figure 8 illustrates the recovery function trajectories ($r_i(t)$) quantifying residual tissue stress during the shift. All segments demonstrated incomplete recovery during breaks (values remained >0.4), with shoulders exhibiting the slowest recovery dynamics ($r_{\text{shoulder}}(t) = 0.88\text{--}0.92$ during work periods). The trunk showed the poorest recovery capacity ($r_{\text{trunk}}(t)$ reduced from 0.91 to only 0.65 during breaks), directly contributing to its accelerated post-break accumulation. These patterns align with the muscle oxygenation kinetics literature [8,15], in which larger muscle groups (trunk) exhibit slower recovery than smaller segments (neck) under sustained loading conditions.

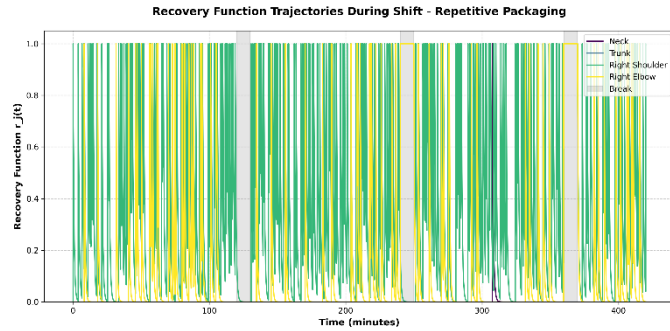


Figure 8. Recovery function trajectories for critical segments during Task 1.

Vertical gray bands indicate break intervals. Higher values indicate greater residual tissue stress. The trunk demonstrates the slowest recovery dynamics (values remain >0.65 during breaks), directly contributing to progressive accumulation of tissue stress across work intervals, a critical physiological mechanism underlying cumulative MSD risk.

C. Segment-Level Exposure Metrics and Risk Reclassification

Table II synthesizes segment-specific exposure metrics derived from the simulation. The trunk demonstrated the highest cumulative burden (971.75 exposure-minutes) due to sustained moderate-to-high flexion without adequate recovery periods, while the neck showed the most effective recovery during breaks ($RR = 0.42$).

Critically, despite moderate instantaneous REBA scores (5.13 = Moderate Risk p), the cumulative Risk Score reached 3,526.38 exposure-minutes, classified as Very High Risk per our threshold framework ($>1,000$ exposure-minutes). This represents a 689% increase in risk classification (Moderate $\rightarrow$ Very High) when moving from instantaneous to cumulative assessment.

Table II

SEGMENT-LEVEL CUMULATIVE EXPOSURE METRICS FOR TASK 1

	FINAL EXPOSURE (EXP-MIN)	Avg. Score	Time in Max Score*	Recovery Ratio*
Neck	736.86	1.95	398.1 min (94.8%)	0.42
Trunk	971.75	2.50	211.7 min (50.4%)	0.28
Right shoulder	498.12	1.57	9.1 min (2.2%)	0.38
Right elbow	438.82	1.50	210.7 min (50.2%)	0.51
CRS	3,526.38	/	/	/

*Time in Max Score: Cumulative duration spent at the highest observed risk score for that segment

*Recovery ratio = exposure slope reduction during breaks

D. Hazardous Pattern Detection

Sliding window analysis detected task-specific hazardous signatures with critical segment-specific variations:

- Trunk: Detected all three hazardous patterns:
 - Pattern 1 (Sustained posture >120 s): $\checkmark$ YES (consecutive Score 3 episodes exceeding 180 s)
 - Pattern 2 (Insufficient recovery $RR < 0.3$): $\checkmark$ YES ($RR = 0.28 < 0.3$ threshold)
 - Pattern 3 (High repetition + low variability): $\checkmark$ YES ($\sigma_\theta = 3.8^\circ$ across 94% of cycles)

This triad of hazardous signatures in the trunk segment represents the highest-risk biomechanical profile observed in this study, directly implicating trunk flexion as the primary driver of cumulative MSD risk in Task 1.

- Neck and Right Elbow: Detected only Pattern 3 (High repetition + low variability), with $\sigma_\theta < 5^\circ$ across $>90\%$ of cycles, confirming the highly repetitive nature of Task 1 without sustained extreme postures.
- Right Shoulder: No hazardous patterns detected beyond brief Score 3 excursions (<15 s duration), indicating lower cumulative risk contribution despite occasional high-risk postures.

E. Temporal Dynamics of Cumulative Risk Accumulation

The most significant finding emerged from the temporal evolution of cumulative exposure (Figure 3). The CRS exceeded the Very High-Risk threshold at minute 218, despite never reaching REBA Scores ≥ 8 , demonstrating that snapshot assessments fundamentally underestimate MSD risk in repetitive tasks. This discrepancy arises because:

- Repetitive moderate-to-high-risk postures (Scores 2-3) generated substantial cumulative burden through sustained exposure without adequate recovery
- Insufficient recovery during breaks (trunk $RR = 0.28 < 0.3$ threshold) permitted progressive residual stress accumulation across work intervals
- Post-break acceleration (trunk slope increase 22%) indicated progressive tissue fatigue despite scheduled rest periods

This finding validates our core hypothesis: temporal integration of exposure is essential for accurate MSD risk prediction in repetitive industrial tasks. The framework reveals hazardous patterns invisible to conventional assessments, specifically, the insidious nature of cumulative loading from repetitive moderate-risk postures combined with inadequate recovery dynamics.

IV. DISCUSSION

A. CUMULATIVE EXPOSURE REVEALS CRITICAL RISK PATTERNS INVISIBLE TO SNAPSHOT ASSESSMENTS

This pilot study demonstrates that temporal integration of segment-specific ergonomic scores fundamentally transforms risk assessment into repetitive industrial tasks. The most salient finding is the dramatic reclassification of Task 1 risk from

Moderate (instantaneous REBA 5.13) to Very High (CRS 3,526.38 exposure-minutes), a 689% increase in risk severity quantification. Critically, this reclassification directly implicates the anatomical segments corresponding to the facility's documented injury epidemiology: trunk (lumbar) and shoulders, the precise regions affected by MSDs requiring >15 days of sick leave in this chocolate manufacturing facility.

The trunk emerged as the primary risk driver (971.75 exposure-minutes), spending 50.4% of the shift duration in REBA Score 3 (>20° flexion) with insufficient recovery ($RR = 0.28 < 0.3$ threshold). This aligns precisely with cumulative trauma pathophysiology: when tissue loading exceeds metabolic recovery capacity over time, micro-damage accumulates in spinal structures [7]. The right shoulder's substantial burden (498.12 exposure-minutes) with sustained elevation (20°–45°) and brief excursions into Score 3 (>45°) directly corresponds to rotator cuff strain mechanisms documented in repetitive overhead tasks [6]. These findings are not theoretical, they reflect the biomechanical origins of the facility's actual MSD burden, where lumbar and shoulder disorders dominate lost-time injury statistics.

The neck's substantial burden (736.32 exposure-minutes), despite never exceeding Score 2, further challenges conventional ergonomic paradigms. Score 2 represents sustained non-neutral posture (>20° flexion or rotation), the minimum risk threshold requiring intervention within days to weeks [2], not a benign state. Our framework quantifies how the 94.8% shift in duration in this configuration generates significant cumulative loading, a phenomenon entirely masked by snapshot assessments that report identical peak scores for brief versus sustained exposures.

B. Physiological Mechanisms Underlying Cumulative Risk Accumulation

The recovery function trajectories ($r_i(t)$) provide direct physiological evidence of tissue stress dynamics corresponding to documented injury mechanisms. The trunk's slower recovery (r_{trunk} decayed to 0.65 during breaks vs. neck to 0.48) aligns with muscle oxygenation kinetics literature [8,15], where larger muscle groups exhibit prolonged metabolite clearance due to greater tissue mass and vascular constraints. This physiological reality explains the trunk's accelerated post-break accumulation (slope increase 22%) and detection of all three hazardous patterns: sustained posture (>120 s), insufficient recovery ($RR < 0.3$), and high repetition ($\sigma_0 = 3.8^\circ$). These patterns match EMG-validated fatigue thresholds [23,30] and directly correspond to the pathophysiological mechanisms underlying lumbar disc degeneration and facet joint overload observed in repetitive flexion tasks [7].

Similarly, the shoulder's sustained moderate elevation (20°–45°) with insufficient recovery periods generates cumulative supraspinatus tendon loading, precisely the mechanism implicated in rotator cuff tendinopathy requiring extended sick leave [6]. The absence of extreme postures (>90° abduction) does not preclude injury risk; rather, duration in moderate-risk configurations without adequate recovery drives tissue micro-damage accumulation. This explains why facility injury records show shoulder disorders emerging from repetitive packaging tasks despite moderate instantaneous scores, a phenomenon invisible to snapshot assessments but captured by our temporal integration framework.

C. Temporal Integration as Essential Ergonomic Metric

This study establishes temporal integration as a non-negotiable component of accurate ergonomic risk assessment for facilities with documented MSD burdens. While prior vision-based systems successfully automated instantaneous scoring [1,21,22], they perpetuated the "snapshot fallacy" by discarding 99.9% of temporal information inherent in high-frequency skeletal data streams. Our framework bridges the critical gap between observational tools (RULA/REBA) and cumulative models (Strain Index, OCRA) by:

- Preserving segment-level anatomical specificity lost in global scores—critical when facility injury data implicates specific segments (lumbar/trunk and shoulders)
- Implementing physiologically grounded recovery dynamics (exponential decay per muscle oxygenation kinetics)
- Detecting hazardous patterns validated against EMG fatigue thresholds and aligned with documented injury mechanisms
- Quantifying exposure in clinically interpretable units (exposure-minutes) directly linked to tissue loading duration

In contrast to forecasting approaches that predict future postures [26,27], our methodology operates exclusively on observed historical exposure, aligning with the physiological reality that MSDs develop from accumulated loading, not predicted states [27]. This distinction is critical: pattern recognition (detecting sustained trunk flexion >120 s without recovery) is biomechanically valid; posture prediction is not. By transforming vision-based systems from reactive posture classifiers into predictive exposure quantifiers, this framework fulfills the preventive vision of Industry 4.0 occupational health [40], specifically targeting the anatomical segments responsible for the facility's documented lost-time injuries.

D. Theoretical Implications for Cumulative Trauma Pathogenesis

Our findings provide empirical validation for cumulative trauma theory by quantifying the temporal dynamics of tissue loading previously accessible only through invasive physiological measurements. The CRS threshold crossing at minute 218, despite never reaching REBA Scores ≥ 8 , demonstrates that MSD risk accumulates through progressive micro-damage when exposure duration exceeds metabolic recovery capacity [7]. This reconciles epidemiological observations (e.g., workers developing lumbar/shoulder disorders without extreme postures [23]) with biomechanical theory: risk derives from integrated exposure (severity $\times$ duration $\times$ frequency $\div$ recovery) rather than peak severity alone.

The trunk's triad of hazardous patterns (sustained posture, insufficient recovery, high repetition) represents a pathophysiological "perfect storm" where tissue loading exceeds clearance capacity across multiple dimensions simultaneously. This explains why lumbar disorders dominate MSD epidemiology in repetitive industrial tasks [4] despite moderate peak scores; cumulative exposure metrics capture the integrated biomechanical insult that snapshot assessments miss. Critically, our framework operationalizes this theory into a quantifiable metric targeting the exact anatomical segments responsible for the facility's documented injury burden (>15 days sick leave for lumbar/shoulder trauma). This alignment between exposure metrics and actual injury patterns provides compelling evidence that temporal integration captures clinically relevant risk dimensions invisible to conventional assessments.

CONCLUSION

This pilot study establishes a methodologically rigorous framework for quantifying cumulative postural exposure through temporal integration of segment-specific ergonomic scores. Applied to Task 1 (repetitive packaging) within a chocolate manufacturing facility characterized by a documented burden of shoulder and lumbar musculoskeletal disorders requiring extended medical leave (>15 days), the framework revealed critical risk patterns undetectable by conventional snapshot assessments. Despite a moderate mean instantaneous REBA score of 5.13 (Moderate Risk classification), cumulative exposure metrics reclassified overall risk as Very High (Cumulative Risk Score = 3,526.38 exposure-minutes). Critically, the anatomical segments exhibiting the highest cumulative burden, the trunk (971.75 exposure-minutes) and right shoulder (498.12 exposure-minutes), align precisely with the facility's documented injury epidemiology. This congruence provides empirical validation that temporal integration captures clinically relevant risk dimensions directly associated with occupational health outcomes.

The framework's methodological contribution resides in recovery-weighted integration of segment scores, which quantifies how sustained exposure to non-neutral postures (REBA Scores 2–3) without adequate recovery generates substantial cumulative loading. The trunk segment demonstrated a confluence of hazardous patterns: sustained non-neutral posture (>120 seconds), insufficient recovery ratio ($RR = 0.28 < 0.3$ threshold), and high repetition (angular variability $\sigma_\theta = 3.8^\circ$). This triad represents a pathophysiological profile consistent with cumulative trauma mechanisms, wherein tissue micro-damage accumulates when loading exceeds metabolic recovery capacity. Similarly, the shoulder's substantial burden despite minimal time in high-risk postures (2.2% in Score 3) illustrates how duration in moderate-risk configurations, not peak severity, drives cumulative tissue stress. These findings empirically substantiate that MSD risk in repetitive industrial tasks is predominantly governed by exposure duration and recovery dynamics rather than instantaneous posture severity.

This investigation acknowledges three methodological constraints. First, validation was confined to a single repetitive task (Task 1) with stable kinematic patterns (cycle time ≈ 30 seconds). While this focused design is methodologically justified for proof-of-concept in highly repetitive operations, generalizability to heterogeneous tasks requires multi-task validation. The minor discrepancy in simulated versus observed trunk Score 3 prevalence (50.4% versus 42%) reflects expected stochastic variation in Markov chain modeling of repetitive tasks and does not compromise physiological plausibility. Second, full-shift exposure trajectories were extrapolated from a 30-second empirical recording. Although validated for tasks with low intra-cycle variability ($\sigma_\theta < 5^\circ$ across 92% of cycles), this approach cannot capture infrequent high-risk postures occurring outside the sampled window. Third, validation relied on physiological plausibility (recovery dynamics aligned with established muscle oxygenation kinetics), pattern consistency with electromyography-validated fatigue thresholds, and congruence with facility injury records. Direct correlation with prospective MSD incidence or longitudinal discomfort metrics was not established.

These limitations delineate a structured validation pathway. Short-term implementation will deploy the framework across multiple operators and tasks using direct long-duration recordings to validate simulation fidelity and implement real-time alerts for hazardous patterns (e.g., sustained trunk flexion >120 seconds without ≥ 30 seconds recovery). Medium-term efforts will integrate worker-reported discomfort metrics to

establish exposure-discomfort correlations and refine recovery parameters using individual physiological characteristics where ethically feasible. Long-term objectives include a prospective cohort study correlating cumulative exposure metrics with MSD incidence over 24 months, specifically targeting the lumbar and shoulder segments implicated in the facility's injury burden.

By transforming ergonomic monitoring from reactive posture classification to predictive exposure quantification, this framework enables targeted interventions at the source of cumulative loading. Prioritizing workstation redesign to minimize sustained trunk flexion, implementing scheduled micro-breaks to elevate recovery ratios above critical thresholds, and adopting task rotation protocols offer a scientifically grounded strategy to mitigate documented shoulder and lumbar disorder burdens. This approach embodies the preventive paradigm of Industry 4.0 occupational health: shifting focus from symptom management to proactive risk mitigation through physiologically informed exposure quantification. Future validation across diverse industrial contexts will determine the framework's broader applicability while preserving its core contribution, the temporal dimension as non-negotiable in accurate MSD risk assessment.

ACKNOWLEDGMENT

The authors gratefully acknowledge the operator who participated in this pilot study and the management of SARL Confi Choc El-Madher (Batna, Algeria) for granting research access to their production facility and supporting evidence-based ergonomic initiatives. This collaboration exemplifies the essential partnership between industry and academia required for meaningful advancement in occupational health.

REFERENCES

- [1] Hignett S, McAtamney L. Rapid entire body assessment (REBA). *Appl Ergon*. 2000;31(2):201–205.
- [2] McAtamney L, Nigel Corlett E. RULA: a survey method for the investigation of work-related upper limb disorders. *Appl Ergon*. 1993;24(2):91–99.
- [3] LI G, BUCKLE P. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics* [Internet]. 1999;42(5):674–695.
- [4] Marras WS, Lavender SA, Leurgans SE, et al. Biomechanical risk factors for occupationally related low back disorders. *Ergonomics*. 1995;38(2):377–410.
- [5] Svendsen SW, Bonde JP, Mathiassen SE, et al. Work related shoulder disorders: quantitative exposure-response relations with reference to arm posture. *Occup Environ Med*. 2004;61(10):844–853.
- [6] Punnett L, Wegman DH. Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology* [Internet]. 2004;14(1):13–23.
- [7] Takala E-P, Pehkonen I, Forsman M, et al. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand J Work Environ Health*. 2010;36(1):3–24.
- [8] Vedsted P, Blangsted AK, Søgaard K, et al. Muscle tissue oxygenation, pressure, electrical, and mechanical responses during dynamic and static voluntary contractions. *Eur J Appl Physiol*. 2006;96(2):165–177.
- [9] Greggi C, Visconti VV, Albanese M, et al. Work-Related Musculoskeletal Disorders: A Systematic Review and Meta-Analysis. *J Clin Med*. 2024;13(13).
- [10] Moore JS, Garg A. The Strain Index: a proposed method to analyze jobs for risk of distal upper extremity disorders. *Am Ind Hyg Assoc J*. 1995;56(5):443–458.
- [11] Occhipinti E. OCRA: a concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics*. 1998;41(9):1290–1311.
- [12] Waters TR, Putz-Anderson V, Garg A, et al. Revised NIOSH equation for the design and evaluation of manual lifting tasks. *Ergonomics*. 1993;36(7):749–776.
- [13] Garg A, Moore JS, Kapellusch JM. The strain index to analyze jobs for risk of distal upper extremity disorders: Model validation. 2007 IEEE International Conference on Industrial Engineering and Engineering Management [Internet]. 2007;497–499.

- [14] David GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup Med (Lond)*. 2005;55(3):190–199.
- [15] Bernard B. Musculoskeletal Disorders and Workplace Factors : A Critical Review of Epidemiologic Evidence for WRMDs [Internet]. NIOSH Publication No. 97-141. 1997 [cited 2024 Jan 25]. Available from: <https://stacks.cdc.gov/view/cdc/21745%0Ahttp://www.cdc.gov/niosh>.
- [16] De Kok J, Vroonhof Paul, Snijders Jacqueline, et al. Work-related musculoskeletal disorders : prevalence, costs and demographics in the EU [Internet]. European Agency for Safety and Health at Work. 2019. Available from: <https://osha.europa.eu/es/publications/msds-facts-and-figures-overview-prevalence-costs-and-demographics-msds-europe/view>.
- [17] Cao Z, Hidalgo G, Simon T, et al. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans Pattern Anal Mach Intell*. 2021;43(1):172–186.
- [18] Sun K, Xiao B, Liu D, et al. Deep high-resolution (HR-NET) . Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [Internet]. 2019;2019-June:5686–5696.
- [19] Fernández MM, Bajo J, Vargas SM, et al. Vision-Based Ergonomic Risk Estimation: Deep-Learning Strategies BT - Soft Computing Applications. In: Balas VE, Jain LC, Balas MM, et al., editors. Cham: Springer International Publishing; 2023. p. 581–589.
- [20] Agostinelli T, Generosi A, Ceccacci S, et al. Validation of computer vision-based ergonomic risk assessment tools for real manufacturing environments. *Sci Rep*. 2024;14(1):1–19.
- [21] Andersen LL, Vinstrup J, Sundstrup E, et al. Combined ergonomic exposures and development of musculoskeletal pain in the general working population: A prospective cohort study. *Scand J Work Environ Health*. 2021;47(4):287–295.
- [22] Benharkat NEH, Bentaalla Kaced S, Chakhrit A, et al. Automatic real-time ergonomic posture assessment using digital models: a case study of manual handling tasks. *International Journal of Advanced Manufacturing Technology*. 2025;963–980.
- [23] Porta M, Pau M, Orrù PF, et al. Trunk Flexion Monitoring among Warehouse Workers Using a Single Inertial Sensor and the Influence of Different Sampling Durations. *Int J Environ Res Public Health*. 2020;17(19).
- [24] Takigami S, Inui A, Mifune Y, et al. Estimation of Shoulder Joint Rotation Angle Using Tablet Device and Pose Estimation Artificial Intelligence Model. *Sensors*. 2024. p. 2912.
- [25] Melchior C, Zanini RR, Guerra RR, et al. Forecasting Brazilian mortality rates due to occupational accidents using autoregressive moving average approaches. *Int J Forecast [Internet]*. 2021;37(2):825–837.
- [26] Carrara F, Elias P, Sedmidubsky J, et al. LSTM-based real-time action detection and prediction in human motion streams. *Multimed Tools Appl [Internet]*. 2019;78(19):27309–27331.
- [27] Yadav M, Verma PK, Ansari S. Investigation of the Role of Machine Learning and Deep Learning in Improving Clinical Decision Making for Musculoskeletal Rehabilitation. *Advances in Distributed Computing and Artificial Intelligence Journal*. 2024;13:1–14.
- [28] Abdel-Malek Daniel M, Foley Ryan C A, Wakeely Fahima, et al. Exploring Localized Muscle Fatigue Responses at Current Upper-Extremity Ergonomics Threshold Limit Values. *Hum Factors [Internet]*. 2022;64(2):385–400.
- [29] Tao F, Qi Q, Wang L, et al. Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. *Engineering*. 2019;5(4):653–661.
- [30] Tao F, Cheng J, Qi Q, et al. Digital twin-driven product design , manufacturing and service with big data. 2018;3563–3576.
- [31] Benharkat NEH, Aslan I. Social-technical hazards and their correlations within the framework of digital transformation. *Organizational Behavior in the Digital World*. 2023.
- [32] Benharkat NEH, Bentaalla-Kaced S, Chergui A. The influence of manufacturing digitisation on safety management and reliability engineering. *International Journal of Reliability and Safety*. 2023;17(2).
- [33] Greco A, Caterino M, Fera M, et al. Digital twin for monitoring ergonomics during manufacturing production. *Applied Sciences (Switzerland)*. 2020;10(21):1–20.
- [34] Tao F, Sui F, Liu A, et al. Digital twin-driven product design framework. *Int J Prod Res [Internet]*. 2019;57(12):3935–3953.
- [35] Jacobs J V, Hettinger LJ, Huang Y-H, et al. Employee acceptance of wearable technology in the workplace. *Appl Ergon [Internet]*. 2019;78:148–156.
- [36] Jiang L. A digital twin-based biomechanical and psychosocial coupling framework for university sports dance training and evaluation. *Discover Artificial Intelligence [Internet]*. 2025;6(1):52.
- [37] Johnen L, Schaub M, Mertens A, et al. Can cumulative loading estimates be used to assess the collective occupational risk of MSD? Evaluation of calculation methods for spinal cumulative loading. *Int J Ind Ergon [Internet]*. 2022;92:103361.
- [38] Hoe VCW, Urquhart DM, Kelsall HL, et al. Ergonomic interventions for preventing work-related musculoskeletal disorders of the upper limb and neck among office workers. *Cochrane Database of Systematic Reviews*. 2018;2018(10).
- [39] Varas M, Chang L, Garcia JC, et al. Risk Assessment of Musculoskeletal Disorders Using Artificial Intelligence. *E3S Web of Conferences*. 2024;532.
- [40] Soares CO, Pereira BF, Pereira Gomes MV, et al. Preventive factors against work-related musculoskeletal disorders: Narrative review. *Revista Brasileira de Medicina do Trabalho*. 2019;17(3):415–430.
- [41] Robertson MM, Ciriello VM, Garabet AM. Office ergonomics training and a sit-stand workstation: Effects on musculoskeletal and visual symptoms and performance of office workers. *Appl Ergon [Internet]*. 2013;44(1):73–85.
- [42] Kishor Annanth V, Abinash M, Rao LB. Intelligent manufacturing in the context of industry 4.0: A case study of siemens industry. *J Phys Conf Ser [Internet]*. 2021;1969(1):12019.
- [43] Memar Ardestani M, Yan H. Noise Reduction in Human Motion-Captured Signals for Computer Animation based on B-Spline Filtering. *Sensors (Basel)*. 2022;22(12).
- [44] Musculoskeletal Disorders and the Workplace. National Research Council (US) and Institute of Medicine (US) Panel on Musculoskeletal Disorders and the Workplace. Washington (DC): National Academies Press (US); 2001.



Nour El Houda Benharkat is a PhD student at the National Polytechnic School of Algiers, specializing in Process Safety and Industrial Risk Control within the Industrial and Environmental Risk Management Department. She obtained her Master's degree in Health and Industrial Safety, with a focus on Industrial Risk Control, from the National Institute of Health and Industrial Safety, University of Hadj Lakhdar Batna, Algeria. Her research interests include safety management, reliability engineering, and industrial risk management.



Souad Bentaalla-Kaced completed her higher education at the National Polytechnic School of Algiers, earning her Engineer, Magister, and Doctorate degrees in Mines and Industries. She has held various professional positions, including Operations Engineer, Exploration Engineer, Operations Department Manager, and Project Manager, and is currently a Lecturer at the National Polytechnic School. She has participated in numerous national and international research projects and has demonstrated strong capabilities both in independent work and collaborative environments. Her current research interests focus on Artificial Intelligence and Emotional Intelligence.



Abdelmalek Chergui received his master's and PhD degrees in chemistry and environment from the National Polytechnic School of Algiers. Currently, he is the Director of Education, Diplomas, and Continuing Education at the National Polytechnic School of Algiers. Also, he is a Professor in the Industrial and Environmental Risk Management Department at the National Polytechnic School. His research interests include water treatment and purification, valorization of Agri-Resources products, air pollution, safety, and reliability.

Adsorption of Ni (II) from Water by Natural Bentonite: Studies on Equilibrium Isotherm, Kinetics

Sabrina Berkani and Faroudja Mohellebi

Abstract—The aim of this work is to investigate the adsorption potential of commercial natural bentonite in the removal of Ni (II) ions from aqueous solution. The characterization studies were performed using X-ray Diffraction, Fourier Transform Infrared Spectrometer, thermal gravimetric analysis, Differential Thermal Analysis and Fluorescence spectroscopy. Capacity and specific surface areas are also given. The results show that our sample is a sodic montmorillonite who contains some impurities like quartz and calcite. We also studied the effect of thermic and acid activation on the adsorption of Ni (II) on bentonite. Batch adsorption studies were carried out with various initial Ni (II) concentrations, adsorbent dosages, pH, and contact time. The kinetic studies showed that the nickel adsorption process followed a pseudo-second order. The adsorption data were fitted to the Freundlich and Langmuir isotherm models. This behaviour is described by a monolayer Langmuir type isotherm. The adsorption process is found to be a second order kinetics.

Keywords—Bentonite, Adsorption, Heavy Metals, Thermic Acid Activation, Freundlich and Langmuir Isotherm.

NOMENCLATURE

AAS	atomic absorption spectrophotometer
BET	Brunauer-Emmett-Teller
CEC	Cation exchange capacity
FTIR	Fourier transform infrared spectroscopy
TGA	Thermogravimetric analysis
XRD	X-Ray diffraction
XRF	X-ray fluorescence

I. INTRODUCTION

Nowadays, the problem of heavy metals became increasingly alarming. A common characteristic of the industrial effluents is related to the fact that they almost always contain toxic metals. The environmental protection forces to limit the contents of these metals to the allowed maximum.

Adsorption is one of the most important physicochemical processes that occur at the solid-liquid and solid-gas interfaces. Adsorption has become a preferred method for removal, recovery and recycling of toxic heavy metals from wastewater [1]. Different conventional and non-conventional type of adsorbents have been tried for removal of various metal ions [2], activated carbon [3], tree fern [4], coconut coir pith [5], sewage sludge [6], silica [7], polymetallic sea nodule [8], modified zeolite [9], etc.

The clay minerals in soil play the role of a natural scavenger by removing and accumulating contaminants in water passing

through the soil. The mechanism of these processes is usually through ion exchange and adsorption. The high specific surface area, chemical and mechanical stability, layered structure, high cation exchange capacity (CEC), and tendency to hold water in the interlayer sites have made the clays excellent adsorbent materials [9].

Pure clay minerals, particularly naturally occurring bentonite has been used as a favoured adsorbent for removal of various metals from aqueous medium. Uses of bentonite for adsorptive removal of Zn(II) [11], natural and Na-exchanged-bentonite for removal of Cr(III), Ni(II), Zn(II), Cu(II) and Cd(II) from water [12] and even kaolinite for Mn(II), Co(II), Ni(II), Cu(II) have yielded significant adsorption [13]. Other important results include adsorption of Cu(II) and Zn(II) on surfactant modified montmorillonite [14].

Bentonites are highly valued for their sorptive properties that could possibly be improved by chemical treatment or by heat treatment, which stem from their high surface area and their tendency to absorb water in the interlayer sites. Acid activation increases the surface area [15], [16], [17]; [18], and modifies the structure of smectites [18], [19], [20], [21].

The most important physical changes in activated smectites is the increase of their specific surface area and the average pore volume [17], [19], [22-25]. The extent of these changes depends on the acid strength and time of treatment [17], [19], [26-28], as well as on temperature [29].

In our work we did two types of treatment to clay sample: A chemical treatment with sulfuric acid at various concentrations, a heat treatment under various temperatures. The choice of these modes of activation was determined by the results of a previous work [29-30], that predict an increased capacity of adsorbing bentonites. Which points to an increase in the adsorption capacity of the bentonite.

II. EXPERIMENTAL

A. Reagents

A stock solution containing 100 mg of Ni(II) per liter was

Manuscript received December 29, 2025; accepted July 6, 2026.

S. Berkani and F. Mohellebi are with the Fossil Energy Valorization Laboratory, Chemical Engineering Department, Ecole Nationale Polytechnique, Algiers, Algeria. (e-mail: s.berkani@hotmail.com; mohellebifaroudja@yahoo.fr).

Digital Object Identifier (DOI): 10.53907/enpesj.v6i1.359

prepared by dissolving $\text{Ni}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ in distilled water and was used to prepare the adsorbate solutions by appropriate dilution.

B. Clay adsorbents

Bentonite was obtained from Maghnia (ROUSSEL) Algiers. It was cleaned, dried and sieved into sizes of 100 μm .

C. Adsorbent characterization

XRD measurement

to «X'PERT PRO MPD» Philips X-ray spectrometer using Cu K α radiations was used for characterizing the adsorbents.

FTIR measurement

Infrared spectrograms were recorded with a Perkin Elmer FT-IR using KBr pallets.

TGA measurement

Thermal analysis was carried out between 20 °C and 1100 °C using an apparatus: derivatographe.

Fluorescence spectroscopy

Fluorescence spectroscopy analysis was performed according to NF P 15-467 standard at CETIM, using the Philips Analytical device.

Specific surface area

Specific surface was calculated by using the method BET, in the field $0.02 \leq P/P_0 \leq 0.35$. The volume of the micropores, the specific surface of the micropores and external surface, mainly the surface of the average pore was determined by the method t-plot. The difference between specific surface BET and external specific surface is indicated as being the specific surface of micropores.

Isotherms of adsorption of N_2 on bentonite samples were obtained on an apparatus TRISTAR 3000 (Micromeritics). Roughly 0.74 g of sample was degassed with 200°C during 24 hours under a high vacuum. An isotherm was then obtained at the temperature of liquid nitrogen (77.3K). The size of the pores is given starting from the phase of adsorption of the isotherm of adsorption of nitrogen (method BJH).

Cation exchange capacity

Method used to determine cation exchange capacity is SCHOLLENBERGER. This method consists in moving all the cations by ammonium a normal acetate solution (pH=7) from which excess is eliminated by a solution of ethanol with 95%. In the second percolation, salt used is the potassium chloride 1N.

D. Adsorption experiments

The adsorption experiments were made by batch technique at room temperature (25°C) with known amount of the adsorbent and 200 mL of metal solution at fixed concentrations of metal. The Erlenmeyer were kept under constant agitation (250 rpm). Afterward, samples were taken out from the shaker at regular contact time intervals, and the clay was separated by filtering. The concentration of remaining nickel in the solution, after

shaking, was analysed by using atomic absorption spectrophotometer (AAS).

The adsorbed nickel amount (q_e) per unit absorbent mass was calculated as follows (Equation 1):

$$q_e = \frac{(C_0 - C_e)}{m} \cdot V \quad (1)$$

Where C_0 is the initial Ni^{2+} concentration, C_e is the concentration of Ni^{2+} at equilibrium time (mg/L), m is the clay mass (mg) and V is the solution volume (L).

The following conditions were maintained for different sets of experiments:

Thermic activation: Clay 5 g/L, $\text{Ni}(\text{II}) = 100$ mg/L, pH =7.66, time = 6h, temperature = 100 to 300 °C;

Acid activation: Clay 5 g/L, $\text{Ni}(\text{II}) = 100$ mg/L, pH =7.66, time = 6h, temperature = 98 °C, $[\text{H}_2\text{SO}_4] = 3$ to 36 % mass;

Kinetics: Clay 5 g/L, $\text{Ni}(\text{II}) = 100$ mg/L, temperature 289 K, pH=7.66, interaction time: from 0 to 180 min;

Effects of pH: Clay 5 g/L, $\text{Ni}(\text{II}) = 100$ mg/L, temperature 289 K, interaction time 180 min, pH 1.47 to 8.6;

Effect of adsorbate amount Clay 5 g/L, temperature 295 K, pH=7.66, interaction time 180 min, $\text{Ni}(\text{II}) = 10$ to 300 mg/l.

III. RESULTS AND DISCUSSION

A. Adsorbent characterization

XRD and FT-IR study

XRD show intense peaks characteristic of montmorillonite, with the values ($d = 4.46\text{\AA}$, $2\theta = 19.91^\circ$), ($d = 3.19\text{\AA}$, $2\theta = 27.77^\circ$) and ($d = 3.76\text{\AA}$, $2\theta = 23.60^\circ$).

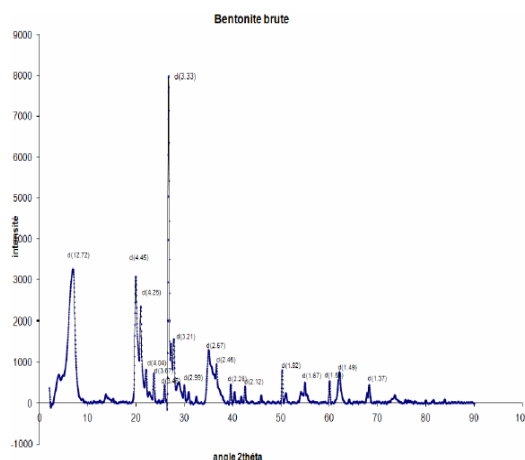


Fig. 1: XRD patterns for bentonite.

A FT-IR spectroscopic study for bentonite is shown in Fig. 2. As we can see there are absorption bands in the range of 3200-3800 cm^{-1} , 1600-1700 cm^{-1} .

The bands in the range of 1600-1700 cm^{-1} are due to lattice OH and bound water stretching vibrations. A strong and sharp band is detected at 1040 cm^{-1} which is related to Si-O stretching vibrations. The weak absorption bands at 528 and

478 cm^{-1} can be attributed to the Si-O-Al(VI) et Si-O-Mg(VI) bending vibrations. The band in the range of 3200-3800 cm^{-1} , with peaks at 3625 and 3440 cm^{-1} characterizing montmorillonite, corresponds to stretching vibrations of OH-.

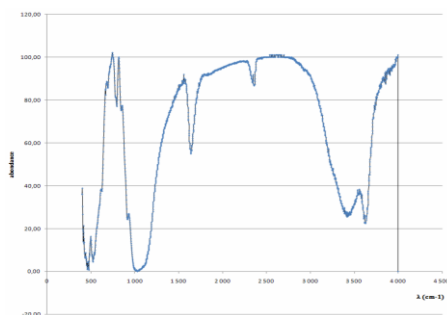


Fig.2: FT-IR (Fourier Transform Infrared spectrophotometer) Spectra of the natural bentonite.

TGA/ATD study

The interpretation of the DTA-TGA curves of bentonite leads to the following results:

- An endothermic peak at 100 $^{\circ}\text{C}$, due to the removal of the hygroscopic water interlayer.
- An endothermic peak between 550 and 650 $^{\circ}\text{C}$ which corresponds to the loss of hydroxyl group from the clay mineral structure (clay dihydroxylation).
- An endothermic peak between 820 and 920 $^{\circ}\text{C}$ shows the formation of a phase at high temperature characteristic of aluminium ferifere smectite.

In addition to these peaks, other peaks appear in the DTA curve of the crude sample corresponding to the transformation of the quartz, at 650 $^{\circ}\text{C}$.

Curve of TGA allow us to follow the weight loss of bentonite. These curves show three losses of weight:

- The first loss of weight occurs between 20 and 180 $^{\circ}\text{C}$. These transformations are due to the removal of adsorbed and interlayer water from the clay mineral.
- The second loss of weight occurs between 400 and 550 $^{\circ}\text{C}$ due the removal of water composition from the clay minerals. The third loss of weight between 600 and 700 $^{\circ}\text{C}$. This loss can be attributed to the dihydroxylation of bentonite.

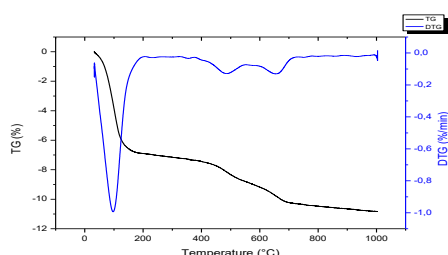


Fig. 3: DTA-TGA curve of bentonite.

Table. 1

CONTENTS OF MAJOR ELEMENTS IN SAMPLE BENTONITE

Compounds	SiO ₂	Al ₂ O ₃	CaO	MgO	MnO
Content (%)	48.09	12.80	0.52	8.22	0.21
Compounds	Fe ₂ O ₃	TiO ₂	P ₂ O ₅	Na ₂ O	K ₂ O
Content (%)	12.60	0.36	0.02	2.02	0.83

Fluorescence spectroscopy study

X-ray fluorescence (XRF) of sedimented fine (less than 2 μm) Maghnia bentonite shows a significant percentage of silicon oxide (SiO₂) and aluminium oxide (Al₂O₃), and a low percentage K₂O (Table 1) which approves that it is a bentonite which belongs to the families of montmorillonite.

Surface area

We plot a graph Q_m the amount adsorbed as a function of relative pressure P/P_0 . The shape of the curve is shown in Fig. 4.

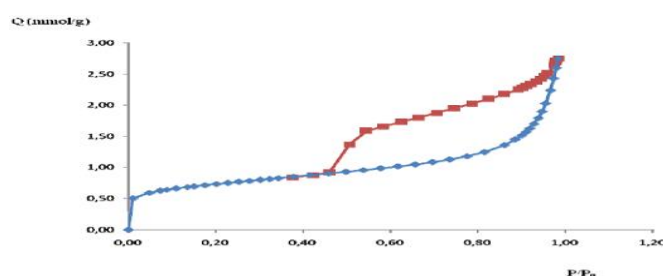


Fig. 4: Adsorption and desorption isotherms of nitrogen vapor at 77°K.

The nitrogen adsorption isotherms at 77°K for our clay (Fig. 4) are Type IV isotherms [31], the hysteresis effect on desorption being an indication of the presence of pores in the intermediate size range.

The results obtained show that our material is a mesoporous because we have a low microporous volume of $24.3 \times 10^{-3} \text{ cm}^3/\text{g}$ and a specific surface equal to $56.44 \text{ m}^2/\text{g}$.

Cation exchange capacity (CEC)

For the divalent cations, magnesium is the most dominating element (46.0 meq/100g). It is four times quantity of calcium (11.0 meq/100g). For monovalent cations, the most dominating element is sodium (36.7 meq/100g). Thus, the

Table.2

CATION CHARACTERISTICS AND AMOUNT OF VARIOUS CATION COMPOUNDS IN BENTONITE.

Cation	Sample meq/100g
Magnesium	46.0
Sodium	36.7
Calcium	11.0
Potassium	27.0
Exchangeable cation content (S)	120.7
Cation exchange capacity (T)	122.3

dominating replacement parts in the bentonite sample are Magnesium and Sodium.

Capacity of total exchange (T) is higher than the sum of the exchangeable cations (S). The difference (T-S) is 1.4

meq/100g of dry product (table 2).

This difference (T-S) can be due maybe with:

- The used method (dynamic method),
- Salts employed whose dimensions of molecules are relatively high,
- The probable presence of other cations such as Cu^{++} , Fe^{++} , Mn^{++} , ...etc.

B. Adsorption of Ni^{2+}

Thermic and acid activation

Fig. 5 shows that the activated bentonite in a temperature between 120 and 160°C is the one which represents the biggest capacity of exchange. This increase can be justified by the elimination of a part of the molecules of adsorbed water. This departure of water results the liberation of positions of exchange. The decreasing values of the activity of the clay, which we obtain afterward, would be possibly the direct consequence of the progressive destruction of the crystalline structure of the clay with the rise of temperature. When this material is heated at temperature higher than 160°C, it loses more and more its adsorption properties.

For the rest of our work, it will be economic to use a bentonite dried just at 105-110°C, because there is a small difference between the quantity of nickel fixed on bentonite at 105°C ($Q_e = 18.44 \text{ mg.g}^{-1}$) and that fixed on activated clay in a range of temperature 120-160°C ($Q_e = 18.60\text{-}18.65 \text{ mg.g}^{-1}$). According to the Fig. 5, (18.44 mg.g^{-1}) represents the maximum quantity of nickel adsorbed by natural bentonite. The Ni^{2+} removal efficiency of bentonite was approximately 92 %.

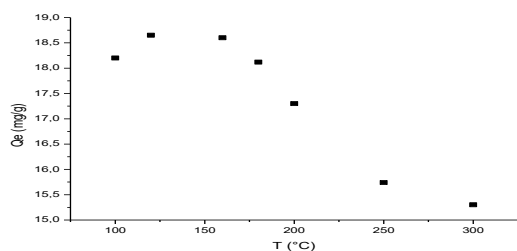


Fig. 5: Removal of Ni^{2+} by thermic activated bentonite at equilibrium.

According to researchers, the acid activation of the clay is due to the development of pores after uniform dissolution of the structure so that it has exactly one silicium atom for one of aluminium atom. When the solution is saturated with silicium, aluminium continues to dissolve and disappear from the network, causing the collapse. This explains the decrease in adsorption capacity of bentonite due to the rapid collapse of the structure after activation.

As we can see (Fig. 5 and Fig.6), the two types of activation of bentonite (chemical and thermic) have a negative impact on the ability of adsorption of our clay. That is why we continue our research with natural Bentonite just dried at 105-110 °C.

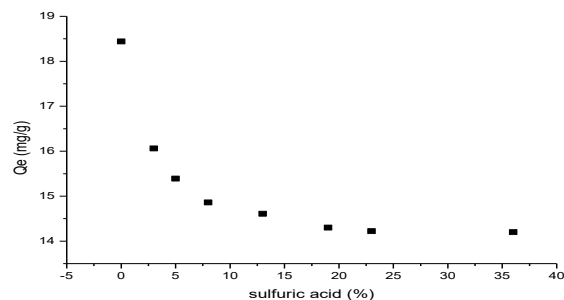


Fig. 6: Removal of Ni^{2+} by acid activated bentonite at equilibrium.

Effect of pH

Fig. 7 shows that adsorption of Ni(II) increase from 67.1% to 76.5% in the pH range of 1.47 to 6.31 under the experimental conditions.

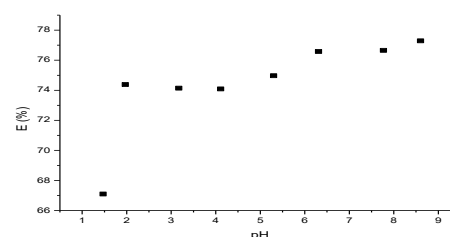


Fig. 7: Influence of pH on adsorption of Ni(II) on bentonite.

Then variation of adsorption remains constant until pH close to 9. At pH = 8.08, we have the rate of the highest elimination of nickel, (79.06%).

When pH is close to 7, the mechanism of adsorption by ionic exchange is supported. At these values of pH, exchangeable cations present on the sites of exchange (i.e. Na^+ , K^+ , Ca^{2+} et Mg^{2+}) are exchanged with the cations $\text{Ni}(\text{II})$ present in solution [32, 33].

If the solution is basic, adsorption by complexation of surface is favoured. The principal mechanism passes from the ionic exchange to the complexation of surface with the increase of pH. When pH increase, the dominant aqueous species of nickel changes into $\text{Ni}(\text{OH})_2$, $\text{Ni}(\text{OH})_3$ et $\text{Ni}(\text{OH})_4^{2-}$ with weak charge, these species have a weaker competitiveness for the ionic exchange [34, 35]. In general, the influence of pH is negligible because it does not improve much the yield. (An increase of 10%).

Effects of adsorbate amount

Fig. 8 shows that the amount adsorbed per unit mass, q_e , increase with the initial concentration of Ni(II), up to 100mg/l quantity adsorbed remain constant. This may be explained by the fact that in the diluted solutions, the mobility of the ions nickel is high, so his interaction with the adsorbent increases.

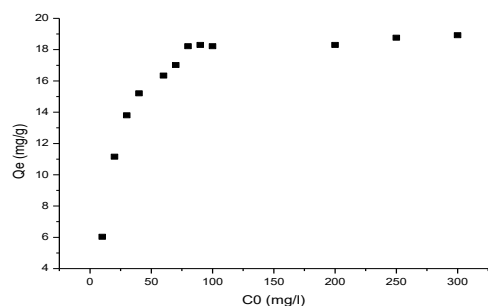


Fig. 8: Effects of Ni(II) concentration for adsorption on bentonite.

According to some authors [36, 37] at low initial metal ion concentration would have first the fixing of the cations on sites of strong chemical reactivity with side surface.

With an increase in Ni(II) concentration, a second site of lower chemical reactivity with stronger surface (side surface) would intervene whereas the sites of strong chemical reactivity would be almost saturated.

Kinetics of Ni(II) adsorption

The curve characteristic of the adsorption of nickel is respectively represented on Fig.9. This kinetics is characterized by a fast fixing; after only 3 minutes, the bentonite absorbed 80% of nickel. Beyond 140 min, the quantity of nickel fixed does not vary.

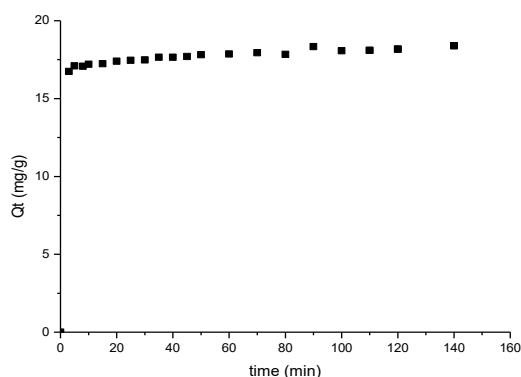


Fig. 9: Effects of interaction time on adsorption of Ni(II) on bentonite (clay 5/L, initial Ni(II) 100 mg/L.).

Different models have been used to test the kinetics of clay-Ni(II) interactions. The pseudo first order curves are obtained by plotting $\log [q_e(q_e - q_t)]$ vs. time. The first order rate constant (obtained from the slopes) equal to $2.53 \cdot 10^{-2} \text{ min}^{-1}$ (Table 3).

However, linearity of pseudo first order plots does not

Table. 3

RATE COEFFICIENTS FOR ADSORPTION OF Ni(II) ON BENTONITE.

Pseudo first order		Second order	
$k_1(\text{min})$	R^2	$K_2(\text{g/mg} \cdot \text{min})$	R^2
$2,53 \cdot 10^{-2}$	0,677	$7,10^{-2}$	0,999

necessarily assure a first order mechanism. The large deviations have led to verifying the kinetics of the interactions further by using the second order equation.

The kinetics of Ni(II) adsorption on clays is very close to second order kinetics.

Adsorption isotherm

The Langmuir plots based on equation 2 [38] :

$$\frac{C_e}{q_e} = \frac{1}{(b \cdot q_m)} + \left(\frac{1}{q_m}\right) C_e^* \quad (2)$$

The Freundlich plots, based on equation 3 [39]:

$$q_e = K_F + C_e^{1/n*} \quad (3)$$

The values of the adsorption coefficients obtained from the plots are given in Table 4.

It arises from the table 4 that it is the equation of Langmuir

Table.4

FREUNDLICH AND LANGMUIR COEFFICIENTS FOR ADSORPTION OF Ni(II) ON

BENTONITE		
Langmuir	$Q_m(\text{mg g}^{-1})$	19,61
	$b (\text{L g}^{-1})$	0,10
	R^2	1,00
Freundlich	$K_F (\text{mg}^{1-1/n} \text{L}^{1/n} \text{g}^{-1})$	5,46
	n	4,03
	R^2	0,74

who models at best the contact solution nickel-clay chosen.

Thermodynamic parameters

Thermodynamic parameters, i.e. heat of adsorption, ΔH^0 and entropy change, ΔS^0 , for the sorption of Ni(II) on bentonite were calculated for each system by using the following equations (Equation 5 and Equation 6)

$$\ln K_d = (\Delta S/R) - \left(\frac{\Delta H}{R * T}\right) \quad (5)$$

$$K_d = (Q_e/C_e) \quad (6)$$

The values of ΔH^0 and ΔS^0 are obtained from the slope and the intercept of each plot. The free energy for the specific adsorption, ΔG^0 , was calculated by using the equation 7.

$$\Delta G^0 = \Delta H^0 - T * \Delta S^0 \quad (7)$$

In general, the phenomenon of adsorption is always accompanied by a thermal process [40, 41] which can be either exothermic ($\Delta H < 0$) or endothermic ($\Delta H > 0$). Measurement of adsorption heat is the main criterion for differentiating chemisorption from physisorption.

The values of the thermodynamic parameters for the sorption of Ni(II) on bentonite are given in Table 5.

The positive value of ΔH (Table 5) show that the sorption of nickel ions on bentonite is an endothermic process. The value of ΔG indicates the spontaneity of the process These results show that sorption of Ni(II) on bentonite is an endothermic process.

According to some authors [42, 43] the sorption reactions of the monovalent cations are exothermic, resulting in a sorption that tends to decrease with temperature. Conversely, the sorption reactions of bi and trivalent cations are endothermic, which therefore leads to an increase in retention when the temperature rises.

The positive value of ΔH confirms that the adsorption of nickel on the clay is an endothermic process. This value is less than 40 K Joule / mole. This indicates that it is a physical adsorption, even the negative value of the free enthalpy given in Table 5 shows that the process of adsorption of nickel on bentonite is a spontaneous process [44].

IV. CONCLUSION

The fixing of Ni(II) on Bentonite increases with pH in the interval considered $1.47 \leq \text{pH} \leq 8.60$. The maximum effect is reached for $\text{pH} = 7$. This clay fixes Ni(II) in a very short time ($t=3$ minutes, 80% of Ni(II) retained). Equilibrium is reached at time $t < 10$ minutes. This kinetics is very good because most of the factors are favourable (speed of agitation, temperature, ray of the particles, etc...). This behaviour is described by a monolayer Langmuir type isotherm. The adsorption process is found to be a second order kinetics. These results show that bentonite can be used effectively for the removal of Ni(II). Considering the abundance and low price of bentonites and their physical and chemical characteristics, they are materials to be efficiently applied in the environmental industry. They can be used in the treatment of mining and/or industrial effluents with metallic contents above the standard values considered by the corresponding legislation.

REFERENCES

- [1] T. W. Chang, M. K. Wang, and C. Lin, "Adsorption of copper in the different sorbent/water ratios of soil systems," *Water, Air, and Soil Pollution*, vol. 138, pp. 199–209, 2002, doi: 10.1023/A:1015551016833.
- [2] V. K. Gupta, M. Gupta, and S. Sharma, "Process development for the removal of lead and chromium from aqueous solutions using red mud—an aluminium industry waste," *Water Research*, vol. 35, no. 5, pp. 1125–1134, 2001.
- [3] R. Leyva-Ramos, L. A. Bernal-Jacome, J. Mendoza-Barron, L. Fuentes-Rubio, and R. M. Guerrero-Coronado, "Adsorption of zinc(II) from an aqueous solution onto activated carbon," *Journal of Hazardous Materials*, vol. 90, pp. 27–38, 2002, doi: 10.1016/S0304-3894(01)00333-8.
- [4] A. Y. S. Ho, C. T. Huang, and H. W. Huang, "Equilibrium sorption isotherm for metal ions on tree fern," *Process Biochemistry*, vol. 37, pp. 1421–1430, 2002, doi: 10.1016/S0032-9592(02)00036-5.
- [5] K. Kadirvelu and C. Namasivayam, "Activated carbon from coconut coirpith as metal adsorbent: Adsorption of Cd(II) from aqueous solution," *Advances in Environmental Research*, vol. 7, pp. 471–478, 2003, doi: 10.1016/S1093-0191(02)00018-7.
- [6] S. C. Pan, C. C. Lin, and D. H. Tseng, "Reusing sewage sludge ash as adsorbent for copper removal from wastewater," *Resources, Conservation and Recycling*, vol. 39, pp. 79–90, 2003, doi: 10.1016/S0921-3449(02)00122-2.
- [7] N. Chiron, R. Guilet, and E. Deydier, "Adsorption of Cu(II) and Pb(II) onto a grafted silica: Isotherms and kinetic models," *Water Research*, vol. 37, pp. 3079–3086, 2003, doi: 10.1016/S0043-1354(03)00156-8.
- [8] S. Maity, S. Chakravarty, S. Bhattacharjee, and B. C. Roy, "A study on arsenic adsorption on polymetallic sea nodule in aqueous medium," *Water Research*, vol. 39, pp. 2579–2590, 2005, doi: 10.1016/j.watres.2005.04.054.
- [9] U. Wingenfelder, B. Nowack, G. Furrer, and R. Schulin, "Adsorption of Pb and Cd by amine-modified zeolite," *Water Research*, vol. 39, pp. 3287–3297, 2005, doi: 10.1016/j.watres.2005.05.017.
- [10] R. Naseem and S. S. Tahir, "Removal of Pb(II) from aqueous/acidic solutions by using bentonite as an adsorbent," *Water Research*, vol. 35, pp. 3982–3989, 2001, doi: 10.1016/S0043-1354(01)00130-0.
- [11] S. Veli and B. Alyüz, "Adsorption of copper and zinc from aqueous solutions by using natural clay," *Journal of Hazardous Materials*, vol. 149, pp. 226–233, 2007, doi: 10.1016/j.jhazmat.2007.04.109.
- [12] E. Alvarez-Ayuso and A. Garcia-Sanchez, "Removal of heavy metals from waste waters by natural and Na-exchanged bentonites," *Clays and Clay Minerals*, vol. 51, pp. 475–480, 2003, doi: 10.1346/CCMN.2003.0510501.
- [13] O. Yavuz, Y. Altunkaynak, and F. Guzel, "Removal of copper, cobalt and manganese from aqueous solution by kaolinite," *Water Research*, vol. 37, pp. 948–952, 2003, doi: 10.1016/S0043-1354(02)00409-8.
- [14] H. Lin and R. S. Juang, "Heavy metal removal from water by sorption using surfactant-modified montmorillonite," *Journal of Hazardous Materials*, vol. 92, pp. 315–326, 2002, doi: 10.1016/S0304-3894(02)00026-2.
- [15] R. E. Grim and G. Kulbicki, "Thermal analysis of clay minerals," *American Mineralogist*, vol. 46, pp. 1329–1333, 1961.
- [16] T. Lan, D. Kaviratna, and T. Pinnavaia, "Clay-reinforced polymer nanocomposites," *Chemistry of Materials*, vol. 6, pp. 573–575, 1994, doi: 10.1021/cm00041a002.
- [17] E. Srasra, F. Bergaya, H. Van Damme, and N. K. Arguib, "Surface properties of an activated bentonite: Decolorization of rapeseed oil," *Applied Clay Science*, vol. 4, pp. 411–421, 1989, doi: 10.1016/0169-1317(89)90019-7.
- [18] J. P. Rupert, W. T. Granquist, and T. J. Pinnavaia, "Catalytic properties of clay minerals," in *Chemistry of Clay and Clay Minerals*, A. C. D. Newman, Ed. London, U.K.: Mineralogical Society, 1987, pp. 275–318.
- [19] C. N. Rhodes and D. R. Brown, "Structural characterization and optimization of acid-treated montmorillonite and high-porosity silica supports for ZnCl₂ alkylation catalyst," *Journal of the Chemical Society, Faraday Transactions*, vol. 88, pp. 2269–2274, 1992, doi: 10.1039/FT9928802269.
- [20] I. Teak, P. Komadel, and D. Muller, "Acid treated montmorillonite: A study by ²⁹Si and ²⁷Al MAS-NMR," *Clay Minerals*, vol. 29, pp. 11–19, 1994, doi: 10.1180/claymin.1994.029.1.02.
- [21] T. Alemdaroglu, G. Akkus, M. Onal, and Y. Sarikaya, "Investigation of the surface acidity of a bentonite modified by acid activation and thermal treatment," *Turkish Journal of Chemistry*, vol. 27, pp. 675–682, 2003.
- [22] G. A. Mills, J. Holmes, and E. B. Cornelius, "Acid activation of some bentonite clays," *Journal of Physical and Colloid Chemistry*, vol. 54, pp. 1170–1185, 1950, doi: 10.1021/j150482a009.
- [23] R. E. Grim, *Applied Clay Mineralogy*. New York, NY, USA: McGraw-Hill, 1962.
- [24] J. Fijal, Z. Klapayta, B. Kwiecinska, J. Zietkiewicz, and M. Zyla, "On the mechanism of acid activation of montmorillonite. II. Changes in morphology and porosity in the light of electron microscopic and adsorption investigations," *Minerals*, vol. 6, pp. 49–57, 1975.
- [25] D. A. Morgan, D. B. Shaw, T. C. Sidebottom, T. C. Soon, and R. S. Taylor, "The function of bleaching earths in the processing of palm, palm kernel and coconut oils," *Journal of the American Oil Chemists' Society*, vol. 62, pp. 292–299, 1985, doi: 10.1007/BF02541394.
- [26] G. A. Kolta, I. Novak, Z. E. T. Samir, and A. E. B. Kamelia, "Evaluation of bleaching capacity of acid-leached Egyptian bentonites," *Journal of Applied Chemistry and Biotechnology*, vol. 26, pp. 355–360, 1975, doi: 10.1002/jctb.5020260152.
- [27] S. C. Kheok and E. E. Lim, "Mechanism of palm oil bleaching by montmorillonite clay activated at various acid concentrations," *Journal of the American Oil Chemists' Society*, vol. 59, pp. 129–131, 1982, doi: 10.1007/BF02662259.
- [28] I. Zaki, M. Abdel-Khalik, and G. M. Habashi, "Acid leaching and consequent pore structure and bleaching capacity modifications of Egyptian clay," *Colloids and Surfaces*, vol. 17, pp. 241–249, 1986, doi: 10.1016/0166-6622(86)80249-9.
- [29] N. K. Madirov, *Bases Théoriques de l'Activation et Mécanismes de l'Action des Sorbants Naturels dans le Processus de Décoloration des Huiles Végétales*. Moscow, Russia, 1973.
- [30] M. Makki and C. Flicoteaux, *Bulletin de la Société Chimique de France*, no. 2, p. 15, 1976.
- [31] S. Brunauer, L. S. Deming, and W. S. Teller, "Adsorption of gases in multimolecular layers," *Journal of the American Chemical Society*, vol. 62, p. 1723, 1940.
- [32] M. Naim, "Hydraulic, diffusion, and retention characteristics of inorganic chemicals in bentonite," Ph.D. dissertation, Dept. Civil and Environmental Engineering, Univ. South Florida, Tampa, FL, USA, 2004.
- [33] K. E. Roehl and K. Czurda, "Diffusion and solid speciation of Cd and Pb in clay liners," *Applied Clay Science*, vol. 12, pp. 387–402, 1998.
- [34] G. Echevarria, M. I. Sheppard, and J. L. Morel, "Effect of pH on the sorption of uranium in soils," *Journal of Environmental Radioactivity*, vol. 53, pp. 257–264, 2001, doi: 10.1016/S0265-931X(00)00116-8.
- [35] D. E. Morris, C. J. Chisholm-Brause, M. E. Barr, S. D. Conradson, and P. G. Eller, "Optical spectroscopic studies of the sorption of UO₂²⁺

- species on reference smectites," *Geochimica et Cosmochimica Acta*, vol. 58, pp. 3613–3623, 1994, doi: 10.1016/0016-7037(94)90153-8.
- [36] H. Viallis-Terrisse, "Interaction des silicates de calcium hydratés, principaux constituants du ciment, avec les chlorures d'alcalins: Analogie avec les argiles," Ph.D. dissertation, Univ. Bourgogne, Dijon, France, 2000.
- [37] L. Gorgeon, "Contribution à la modélisation physico-chimique de la rétention de radioéléments à vie longue par des matériaux argileux," Ph.D. dissertation, Univ. Paris VI, Paris, France, 1994.
- [38] I. Langmuir, "The constitution and fundamental properties of solids and liquids," *Journal of the American Chemical Society*, vol. 38, no. 11, pp. 2211–2295, 1916.
- [39] H. M. F. Freundlich, "Über die Adsorption in Lösungen," *Zeitschrift für Physikalische Chemie*, vol. 57, pp. 385–470, 1906.
- [40] G. Rytwo and E. Ruiz-Hitzky, "Enthalpies of adsorption of methylene blue and crystal violet to montmorillonite," *Journal of Thermal Analysis and Calorimetry*, vol. 71, pp. 751–759, 2003, doi: 10.1023/A:1023309806214.
- [41] A. Ramesh, D. J. Lee, and J. W. Wong, "Thermodynamic parameters for adsorption equilibrium of heavy metals and dyes from wastewater with low-cost adsorbents," *Journal of Colloid and Interface Science*, vol. 291, pp. 588–592, 2005, doi: 10.1016/j.jcis.2005.04.084.
- [42] M. H. Bradbury and B. Baeyens, "Modeling the sorption of Mn(II), Co(II), Ni(II), Zn(II), Cd(II), Eu(III), Am(III), Sn(IV), Th(IV), Np(V), and U(VI) on montmorillonite," *Geochimica et Cosmochimica Acta*, vol. 69, pp. 5875–5892, 2005.
- [43] M. H. Bradbury and B. Baeyens, "Modeling the sorption of zinc and Ni on Ca-montmorillonite," *Geochimica et Cosmochimica Acta*, vol. 63, pp. 325–336, 1999, doi: 10.1016/S0016-7037(98)00281-6.
- [44] B. Benguella and A. Yacouta-Nour, "Elimination des colorants acides en solution aqueuse par la bentonite et le kaolin," *Comptes Rendus Chimie*, vol. 12, pp. 762–771, 2008, doi: 10.1016/j.crci.2008.11.008.

Sabrina BERKANI PhD student in the Fossil Energy Valorization Laboratory, Chemical Engineering Department, Algiers- Algeria. She is particularly interested in the fields of environmental pollution, pollution risk assessment, Adsorbent and water contamination, as well as pharmaceuticals and chemicals.

Faroudja Mohellebi is Professor at the department of chemical engineering of the National Polytechnic School, Algiers, Algeria and was chief of department of chemical engineering from February 2009 to December 2021. She is now Director of the Fossil Energy Valorization Laboratory Professor Mohellebi works on the mobility and bioavailability of traditional contaminants like cadmium and nickel and on compounds of emerging interest like drugs. Her work focuses on analysing contaminants, their chemical speciation, environmental fate and potential impacts. She has participated in the training of many students and has proposed many research topics directly related to the depollution of water loaded with emerging products and pesticides. She has numerous collaborations and projects at national and international level. Her scientific production includes publications, reviews, keynotes, and three book chapters.