A Framework For Critical Infrastructure Monitoring Based On Deep Reinforcement Learning Approach

Kefas Yunana
Department of Computer Science
Nasarawa State University
Keffi, Nigeria
kefasyunana@nsuk.edu.ng

Ishaq Oyebisi Oyefolahan
Department of Information Technology
Federal University of Technology
Minna, Nigeria
o.ishaq@futminna.edu.ng

Sulaimon Adebayo Bashir
Department of Computer Science
Federal University of Technology
Minna, Nigeria
bashirsulaimon@futminna.edu.ng

Abstract—**Critical Infrastructure (CI) are nowadays linked with IOT devices that communicate data through networks to achieve significant collaboration. With the progress in internet connectivity, IOT has disrupt numerous aspects of CI comprising communication systems, power plants, power grid, gas pipeline, and transportation systems. As a disruptive paradigm, the IOT and Cloud computing utilizing Smart IOT devices equipped with numerous sensors and actuating capabilities play significant roles when deployed in CI surroundings with the aim of monitoring vital observable figures consisting of flow rate, temperature, pressure, and lighting situations. Over the years, oil pipeline infrastructure have been the main economic means for conveying refined oil to assembly and distribution outlets. Though damages to the pipelines in this area by exclusion have influence the normal transport of refined oil to the outlets across the country like Nigeria which has influence the stream of income and damages to the environment. Reinforcement Learning (RL) approach for infrastructure reliability monitoring have receive numerous consideration by researchers denoting that RL centered policy reveals superior operation than regular traditional control systems strategies. Many of the studies utilised mainly algorithms for environment with discrete action and observation spaces unlike others with infinite state space. This study proposed a framework for critical infrastructure monitoring based on Deep Reinforcement Learning (DRL) for oil pipeline network and also developed a pipeline network monitoring (PNM) architecture with expression of the environment dynamics as Markov Decision Process. The sample observation space data and strategy for evaluation of the framework was also presented.**

Keywords: Critical Infrastructure, Monitoring, Framework and Deep Reinforcement Learning

I. INTRODUCTION

Regarded as a connected system, internet of things (IOT) comprises of several connected processing devices, gadgets, apparatus and persons that can communicate data through networks to realised meaningful collaboration. With the advancement in internet connectivity, IOT has disrupt several domain of critical infrastructure ranging from power grid, power plants, airports, transportation systems, gas pipeline and communication systems [1] [2] [3]. Despite its numerous advantages, the IOT network have seen large amount of data been generated by sensor devices which pose challenges for near real time processing and bandwidth issues [4] [5].

Infrastructures such as oil and gas pipeline, airports, health care systems, power plants, water treatment plants, and transportation systems are often regarded as Critical infrastructure (CI) which are considered vital for the smooth running of a nation's economy [6] [7].

The connection of CI, IOT and the rapid development of the Internet are facilitating the deployment of strong and reliable results whereby the breakdown of a specific system thus results into a severe catastrophe [8] [9]. As a disruptive paradigms, IOT and Cloud computing utilising Smart IOT gadgets equipped with several sensors and actuating capabilities play vital role when deployed in CI environment with the aim of monitoring essential visible figures comprising of flow rate, temperature, pressure, and lighting situations [10].

Over the years, oil pipeline infrastructure have been the key economic means for conveying refined oil to assembly and distribution outlets. Though, damages to the pipelines in this area by exclusion have influence the regular movement of refined oil to the outlets across the country like Nigeria which has influence the stream of income and damages to the environment [11].

Reinforcement Learning (RL) approach for infrastructure integrity control have drawn the attention of several researchers over the years specifying that RL centered policy shows superior operation than regular tradition control systems policies. More researches includes the planning of maintenance strategies on an operating oil plant by state action reward state action (SARSA) algorithm to exploit the system accessibility and production efficacy.

The utilization of several autonomous machines for armed trucks was likewise considered were the best maintenance time for individual element to reduce the system interruption by Monte Carlo RL was offered. Several of the studies utilised mainly algorithms for environment with discrete action and observation spaces not like majority of critical infrastructure services with endless state space [12]. This study present a framework for critical infrastructure monitoring based on DRL with sample observation space for the oil pipeline parameters and also formulate the PNM as a Markov Decision Process (MDP). The major contributions of the study is as follows:

i. Developed an architecture for an oil pipeline monitoring Agent based on the DRL paradigm.
ii. Formulated the PNM problem based on DRL framework as an MDP

The study comprises of five segments that starts with introduction, related work, proposed framework and overview of DRL, evaluation approach and lastly the conclusion.

II. Related work

CI such as oil and gas pipelines are critical facilities that needs regular monitoring for topmost functioning and safety for extended duration of time. The necessity for oil and gas transmission from production outlet to the receiving end have led to the rise of pipelines infrastructure manufactured globally with the oil and gas usage rate predicted to rise years ahead [13].

Contemporary society frequently depends on several CI that are substantially interlinked with one another and information distribution from one device to the other. This has resulted in the rise of critical infrastructure numbers thus rendering them susceptible to different dangers including regular operating breakdown [11].

Pipeline Infrastructure for transporting oil and gas is considered to be highly critical for a number of countries across the globe with several aging pipelines influenced by diverse distortions such as vandalism, corrosion and cracks resulting in the failure of the pipeline infrastructure [14] [15]. Several pipeline companies exploits different techniques for monitoring of the infrastructure which ranges from patrol near the Right of Way, monitoring acoustic sounds, actual time transient modeling, fibre optic sensing and pressure wave observation [16][17].

Studies reveal that the projected fatalities due to issues comprising transported fuel products using pipeline of longer distance has is commonly available across the world thereby raising the pipeline failure rates [18] [19]. The failures are normally intentional through vandalisation activities or natural cause like device failure and corrosion which results in financial losses and extreme environmental pollution, especially when the leak is not discovered early [18].

The authors in [20] utilized Deep learning approach for Real time CI protection in scenario of flood event. Even thou the study was able to help understand the dependencies among various CI and the severity of the current situation, it however did not examined a particular critical infrastructure as case study.

The study by [21] centered on an effective failure with inducing generation of Cyber Physical systems using Deep Reinforcement Learning (DRL) approach. The study was able to propose a framework for achieving failure inducing input with no realization of the CPS parameters of history logs. The issues observed was that no analysis of a critical infrastructure identified.

The authors in [22] proposed the use of deep reinforcement learning algorithm for anomaly detection in smart environment. The adopted algorithm used has the disadvantage of been a value based algorithm in comparison with Proximal Policy Optimization (PPO) or twin delay deep deterministic policy gradient (TD3) algorithms that uses neural networks to fit both value and policy functions especially for continuous control task like the oil pipeline.

More so, the study of [23] utilized Deep Reinforcement Learning for Interdependent Healthcare Critical Infrastructure Simulation to achieve Dynamic Varying COVID-19 scenario. Even thou the propose scheme shows promising results with regards to covid-19 case study, it was observed that it is limited to the healthcare sector.

The study by [24] proposed a framework for intelligent and secure framework for cyber physical systems. The study was able to deliberate on part of machine learning for improvement of the cyber physical systems. Although the problem to be solved was in the area of security, there was still no clear explanation and illustration of the proposed framework utilizing the machine learning technique.

The authors in [25] applied Swarm Based Deep Learning and Reinforcement based Q learning for Electroencephalography (EEG) Classification with Sparse Autoencoder. The study was able to propose a technique for efficient classification of epilepsy and schizophrenia from the EEG datasets. One of the major limitations of the proposed system is beside the application domain of the health sector, the use of Q learning has the limitation of handling large state space especially for environment with large state space or observation.

The authors in [26] utilised Deep Reinforcement learning for redundancy alleviation for vehicular Collective awareness Services. While the study was able to obtain the best policy on redundancy alleviation utilizing deep Q network, the domain of application is the Vehicular environment.

The study by [27] utilised DRL for dynamic spectrum access in the Multi-Channel Wireless Local Area Networks. The proposed DQN algorithm was used to address the Spectrum Access problem of a Wireless Local Area Networks. The observed limitation was that beside the applications of the algorithm in the aspect of Local Area network, the adopted DQN algorithm is a value based algorithm that does not clearly enhance the reward system.

III. Proposed Framework and Overview of Deep Reinforcement Learning

The world has over the years witness massive disruption by IOT paradigm such that numerous tasks before thought difficult such as reliable management of assets are now feasible. The significance of pipeline reliability management has led to numerous assets and enterprises in realising IOT on pipeline infrastructure systems with the aim of minimising needless charges. The instant result of IOT application on pipeline infrastructure is a wide monitoring system that can offer reliable, real time information about numerous aspects of the pipeline state.

CI monitoring is fundamentally a successive decision making problem whereby decisions have persistent effects. The monitoring decisions not only influence the present form of the asset, but their consequences remain throughout the infrastructure life duration. RL is an aspect of Artificial Intelligence (AI) involved with data focused modelling and resolving successive decision making demands under ambiguity. It comprises three unique features making it suitable for formulating the pipeline infrastructure monitoring. The features includes learning from historical and real time data and ability to handle delay effects and also interact and learn in a stochastic surrounding [28].

DRL as a branch of artificial intelligence is the mixture of RL and Deep Learning (DL) intended to mimic the thinking process of human's. RL emphasizes on resolving consecutive decision making problems concerning numerous real world problems such as video games, driving, sports and portfolio optimization. The objectives bearing in mind the defined problems includes reaching destination safely, winning the game or reducing a product building amount. This involves taking actions and getting responses from the given environment about how close we are to approaching the define goal. Basically the changes that is often observed in the environment and the various responses gotten often heads to the succeeding action to be taken[29].

The idea of RL can best be understood through the notion of an agent relating with an environment at a given state and taking critical actions under specific policies and getting some positive or negative rewards. The architecture of DRL is shown in Fig. 1.

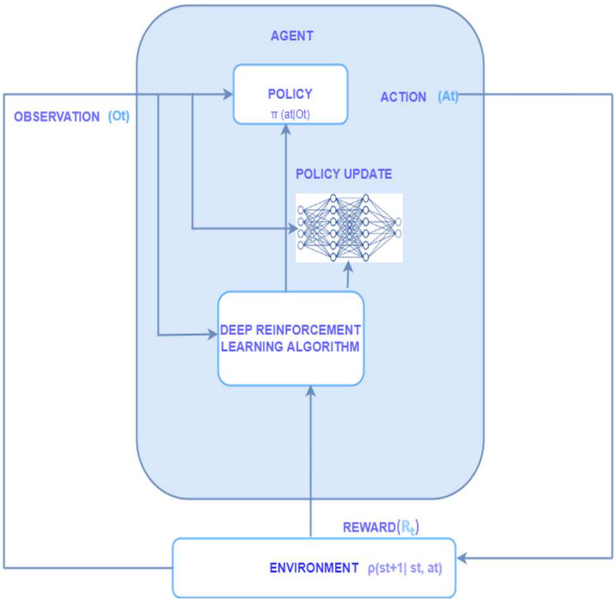

Fig.1. Deep Reinforcement Learning Architecture [30]

The DRL paradigm consists of several parameters that are used in developing various applications which are dependent on the discrete or continuous state space of applications scenario. Key aspect of the RL problem which includes discount factor, rewards and Bellman equation are used for updating Q-Values [29] [30] [31] [32].

The basic form of Bellman equation is given in equation 1 and as a result of the agent environment interaction, the agent gets a reward R(s, a) as shown in equation 2. One of the primary form of RL algorithm is the Q-learning Algorithm that utilizes a lookup table that maps the state and action sets at each time phase t. The basic Q learning at time t is given in equation 3.

$$V(s) = \text{Max } (V(s')) \quad (1)$$
$$V(s) = \text{Max } (R(s, a) + (V(s')) \quad (2)$$
$$Q(s_t, a_t) = r_{t+1} + r_{t+2} + r_{t+3} \ldots\ldots r_{t+n} \quad (3)$$

where $n = 1, 2, 3, 4\ldots\ldots\infty$

To avoid the terms running to infinity and due to the uncertainty of the future rewards as the learning progresses, a discount factor gamma (γ) is introduce with the aim of discounting the future rewards. With the gamma symbol, equation (1) can be rewritten as shown in equation 4. Using a discounted factor less than 1 and raising an exponential term, the number will keep decreasing as the exponential term increases.

$$Q(s_t, a_t) = r_{t+1} + \gamma\, r_{t+2} + \gamma^2\, r_{t+3} \ldots\ldots r_{t+n} \quad (4)$$

Similarly, at time t+1, the Q learning equation is as shown in equation 5. It is worth nothing that part of equation 5 is included in equation 4 when the gamma is multiplied to it as seen in equation 6. Considering equation 6, equation 4 can further be simplified as shown in equation 7. Adjusting the term Q to reflect the aim of taking an action leading to the Max Q value at state t+1 will result in the Q-target equation given in 8.

$$Q(s_{t+1}, a_{t+1}) = r_{t+2} + \gamma\, r_{t+3} + \gamma^2\, r_{t+4} + \ldots\ldots \gamma^n r_{t+(n+2)} \quad (5)$$

where $n = 1, 2, 3, 4\ldots\ldots\infty$

$$Q(s_t, a_t) = r_{t+1} + \gamma[\, r_{t+2} + \gamma\, r_{t+3} + \gamma^2\, r_{t+4} \ \ldots\ldots\ \gamma^{n+1} r_{t+n+2}\,] \quad (6)$$
$$Q(s_t, a_t) = r_{t+1} + \gamma\, Q(s_{t+1}, a_{t+1}) \quad (7)$$
$$Q(s_t, a_t) = r_{t+1} + \gamma \text{ Max } Q(s_{t+1}, a) \quad (8)$$

The Q target equation is the notion used in obtaining the values for a typical Q table and the mathematical form for updating the table is given in equation 9. The key notion regarding the error term in equation 8 is to regulate how fast or slow the Q table update is carried out with a learning rate attached to the error. Adding the learning rate parameter to equation 9 results in equation 10.

$$Q_{New} \longleftarrow Q_{Current} + \text{Error} \quad (9)$$
$$Q_{New} \longleftarrow Q_{Current} + \boldsymbol{\alpha} * \text{Error} \quad (10)$$

With equation 9, a higher learning rate which is a tunable parameter will make the error term to regulate Q table quicker and a lower learning rate will make it slower. Similarly, the error term define the variance between the Q target values and the current Q values as shown in equation 11. Considering equation 8 and substituting equation 11 for the Q target equation results in equation 12 and 13 respectively:

$$Q_{New} \longleftarrow Q_{Current} + \alpha\, [Q \text{ target} - Q_{current}) \quad (11)$$
$$Q_{New} \longleftarrow Q_{Current} + \alpha\, [r_{t+1} + \gamma \text{ Max } Q(s_{t+1}, a) - Q_{current}) \quad (12)$$
$$Q_{New} \longleftarrow Q(s_t, a_t) + \alpha\, [r_{t+1} + \gamma \text{ Max } Q(s_{t+1}, a) - Q(s_t, a_t)) \quad (13)$$

The equation 13 can be re-written as shown in equation 14 to indicate it is an update process with no equal to sign.

The equation 14 is the formal Q learning update equation which is use to update the table as given in table I.

TABLE I. SAMPLE Q TABLE FOR OBSERVATION AND ACTION PAIRS

Observations/States	Actions		
	Do Nothing	Close the Valve	Standby
PR(0),FR(0),TP(0)	Q(s,a)	Q(s,a)	Q(s,a)
PR(1),FR(1),TP(1)	Q(s,a)	Q(s,a)	Q(s,a)
PR(2),FR(2),TP(2)	Q(s,a)	Q(s,a)	Q(s,a)
PR(3),FR(3),TP(3)	Q(s,a)	Q(s,a)	Q(s,a)
‘	‘	‘	‘
‘	‘	‘	‘
‘	‘	‘	‘
PR(n),FR(n),TP(n)			

$$Q(s_t, a_t) \longleftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \text{ Max } Q(s_{t+1}, a) - Q(s_t, a_t)) \quad (14)$$

The problem of continuous state space associated with most real world reinforcement learning has make the Q learning algorithm not practical enough due to the infinite number of observations [29]. This is resolved by introducing deep learning which acts as the function approximation to the reinforcement learning paradigm. For most problems, it is nearly unfeasible to symbolize the Q functions in table form consisting of individual states and action combinations. This has led to the development of neural networks which includes $\boldsymbol{\theta}$ that acts as a constraint for approximating the Q values $Q(s, a; \boldsymbol{\theta}) \approx \mathbf{Q^*(s, a)}$ [29] [31] [32].

A. PNM Problem Design and DRL Formulation

The proposed pipeline network monitoring (PNM) agent consists of an IOT connected system to take sensor readings along the pipeline network which comprises pressure (PR), flow rate (FR) and temperature (TP). The goal of the monitoring agent is to ensure the monitoring parameters are in their optimal range especially during fluid pumping operations to avert leakages. Usually, Pipeline leaks can be reliably modelled using mathematical dynamics of the flow in a gas or oil pipeline. Hence, the leak happening in a pipeline can be reliably projected by observing parameters such as changes in pressure, flow rates and temperature at the upstream and downstream regions of a pipeline following the a leak incidence.

Incidence of leak essentially produces unexpected reduction of the flow characteristics that is proceeded by a limited restoration to its normal rate. Furthermore, the pressure pulsate moves upstream and downstream via the pipeline as wave. Hence, the transient pipeline flow model offers the possibilities for pipeline simulation and modelling.

The goal of the proposed PNM Agent is to ensure prompt response of the state of the pipeline network at any given time on whether the valves should be closed, be in standby mode or remain opened. The system as shown in Fig. 2 comprises of a pipeline network environment, PNM Agent that will be guided by a policy, set of observations or states (PR, FR and TP), reward for taken an action (Do nothing, Close the Valve and Stand by) based on the observations and getting rewarded with a positive or negative number as the case maybe.

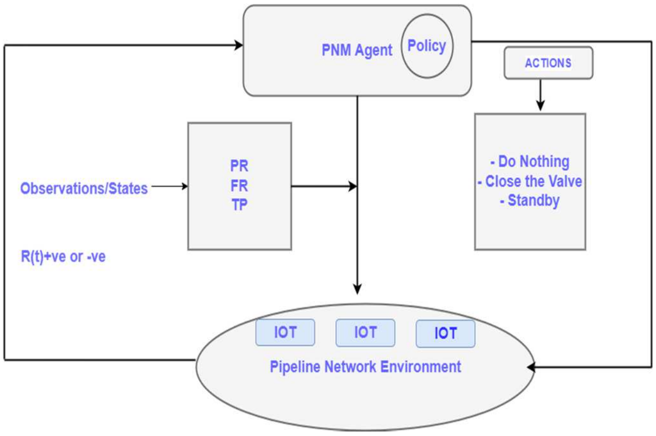

Fig. 2. Proposed PNM Agent for an oil pipeline

In line with the above proposed framework, an MDP in the context of state, action and possible reward for the agent is define.

The state space includes observable parameters generated by the sensors such as pressure (PR), flow rate (FR) and temperature (TP). It is given formally as:

$$S_t = [PR, FR, TP] \in S \qquad (15)$$

Action Space: The action space is defined by available actions that the PNM Agent can take in response of the system state. There are 3 distinct set of actions which includes (Do nothing, Close the Valve and Standby). It is formally given as:

$$A_t = [\text{Do nothing, Close the Valve and Standby}] \qquad (16)$$

Reward: At any point in time in DRL, the agent always gets a reward which can be positive or negative. The goal is to maintain the optimal parameter values at each time step during pumping activities. As such the agent obtains a positive reward for maintaining the optimal values within the specified time frame, negative reward if there are deviations from the optimal values and zero reward otherwise.

The sample observation space which is generated by the environment using sensors of pressure flow, rate and temperature is as shown in table1. The data consist of initial values of 200, 400 and 20 for pressure, flow rate and temperature respectively.

TABLE II. SAMPLE OBSERVATIONS FOR PROPOSED PNM AGENT

Timestamp	**Pressure(PSI)**	**Flow Rate (M^3)**	**Temperature (°C)**
2022-06-24 12:51:10.456032	200.000000	500.000000	20.000000
2022-06-24 12:51:10.456032	357.89473684	684.21052632	21.57894737
2022-06-24 12:51:10.456032	615.78947368	868.42105263	23.15789474
2022-06-24 12:51:10.456032	873.68421053	1052.63157895	24.73684211
2022-06-24 12:51:10.456032	1131.57894737	1236.84210526	26.31578947
2022-06-24 12:51:10.456032	1389.47368421	1421.05263158	27.89473684
2022-06-24 12:51:10.456032	1647.36842105	1605.26315789	29.47368421
2022-06-24 12:51:10.456032	1905.26315789	1789.47368421	31.05263158
2022-06-24 12:51:10.456032	2163.15789474	1973.68421053	32.63157895
2022-06-24 12:51:10.456032	2421.05263158	2157.89473684	34.21052632
2022-06-24 12:51:10.456032	2678.94736842	2342.10526316	35.78947368
2022-06-24 12:51:10.456032	2936.84210526	2526.31578947	37.36842105
2022-06-24 12:51:10.456032	3194.73684211	2710.52631579	38.94736842
2022-06-24 12:51:10.456032	3452.63157895	2894.73684211	40.52631579
2022-06-24 12:51:10.456032	3710.52631579	3078.94736842	42.10526316
2022-06-24 12:51:10.456032	3968.42105263	3263.15789474	43.68421053
2022-06-24 12:51:10.456032	4226.31578947	3447.36842105	45.26315789
2022-06-24 12:51:10.456032	4484.21052632	3631.57894737	46.84210526
2022-06-24 12:51:10.456032	4742.10526316	3815.78947368	48.42105263
2022-06-24 12:51:10.456032	5000.0000000	4000.0000000	50.0000000

IV. EVALUATION APPROACH

The proposed framework will be evaluated considering the different aspect of the monitoring system which includes the environment, an agent and the reward mechanism to incentivize the agent based on the different actions performed at each time step. The environment in reinforcement learning settings represents the changes through which the agent relates with. This involves the environment receiving actions from the PNM agent and in turn outputting the observations considering the modelled dynamics of the environment. The environment then obtains a reward based on the actions taken to measure the agent's performance in achieving the set objectives. The monitoring agent will be created considering one of several standard reinforcement learning algorithms and crafted policies for best performance.

The training and validation will involve an agent interacting with an environment via frequent trial and error procedure with the aim of learning an optimal policy. During the training, the agent adjusts its policy parameters depiction with the aim of maximizing the long term reward. Metrics such as average reward, running mean and cumulative reward will be considered for the selected algorithms and for performance evaluation.

V. Conclusion

CI are nowadays linked with IOT processing devices that can communicate data through networks to realised meaningful collaboration. With the advancement in internet connectivity, IOT has disrupt numerous aspects of CI consisting of power plants, power grid, gas pipeline, communication systems and transportation systems.

The connection of CI, IOT and the rapid development of the Internet are enhancing the deployment of strong and reliable results whereby the failure of a particular system thus results into a severe damages. As a disruptive paradigm, the IOT and Cloud computing utilising Smart IOT gadgets equipped with several sensors and actuating capabilities play key roles when deployed in CI surroundings with the aim of monitoring vital observable figures consisting of flow rate, temperature, pressure, and lighting situations.

Over the years, oil pipeline infrastructure have been the key economic means for conveying refined oil to assembly and distribution outlets. Though, damages to the pipelines in this area by exclusion have influence the regular movement of refined oil to the outlets across the country like Nigeria which has influence the stream of income and damages to the environment.

RL approach for infrastructure reliability monitoring have receive several consideration by researchers over the years denoting that RL centered policy reveals superior operation than regular traditional control systems strategies.
The use of numerous autonomous systems for armed trucks was equally considered were the best maintenance time for specific element to reduce the system interruption by Monte Carlo RL was offered by researchers over the years. Several of the studies utilised mainly algorithms for environment with discrete action and observation spaces unlike others with infinite state space. This study proposed a framework for critical infrastructure monitoring based on DRL with sample observation space for oil pipeline parameters and also formulate the PNM as a Markov Decision Process (MDP). The sample observation data and strategy for the evaluation of the framework was also presented.

References

[1] R. Yu, X. Zhang and M. Zhang, "Smart Home Security Analysis System Based on The Internet of Things," *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, pp. 596-599, doi: 10.1109/ICBAIE52039.2021.9389849.

[2] L. Tang and H. Hu (2020). "Computation Offloading and Resource Allocation for the Internet of Things in Energy-Constrained MEC-Enabled HetNets," *IEEE Access*, 8, pp. 47509–47521, 2020.

[3] Ongfeng Cui, Wei Liu, Pratibha Rani, Melfi Alrasheedi (2021). Internet of Things (IOT) adoption barriers for the circular economy using Pythagorean fuzzy SWARA-CoCoSo decision-making approach in the manufacturing sector, *Technological Forecasting and Social Change*, 171, 120951, https://doi.org/10.1016/j.techfore.2021.120951.

[4] Alaa Omran Almagrabi, A.K. Bashir (2021). A classification-based privacy-preserving decision-making for secure data sharing in Internet of Things assisted applications, *Digital Communications and Networks*, https://doi.org/10.1016/j.dcan.2021.09.003.

[5] D. Dechouniotis, N. Athanasopoulos, and A. Leivadeas (2020). Edge Computing Resource Allocation for Dynamic Networks : The DRUID-NET Vision and Perspective, *Sensors*, 20(2191), pp. 1–18.

[6] L. Wang, J. Zhang, J. Chuan, R. Ma, and A. Fei (2020). "Edge Intelligence for Mission Cognitive Wireless Emergency Networks," *IEEE Wirel. Commun.*, 27(4), pp. 103–109.

[7] M. Chen *et al.*, (2020). "Living with I-Fabric: Smart Living Powered by Intelligent Fabric and Deep Analytics," *IEEE Network.*, 34(5), pp. 156–163.

[8] P. Tedeschi and S. Sciancalepore (2019). "Edge and fog computing in critical infrastructures: Analysis, security threats, and research challenges," *Proc. - 4th IEEE Eur. Symp. Secur. Priv. Work. EUROS PW 2019*, pp. 1–10.

[9] Junaid Farooq and Quanyan Zhu (2021). "Internet of Things-Enabled Systems and Infrastructure," in *Resource Management for On-Demand Mission-Critical Internet of Things Applications*, IEEE, pp.1-8, doi:10.1002/9781119716112.ch1

[10] P. S. Rajulapati, N. Nukavarapu and S. Durbha, "Multi-Agent Deep Reinforcement Learning based Interdependent Critical Infrastructure Simulation Model for Situational Awareness during a Flood Event," *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 6890-6893, doi: 10.1109/IGARSS39084.2020.9323380.

[11] H.A. Umar, M.F. Abdul Khanan, C. Ogbonnaya, M.S. Shiru, A. Ahmad, A.I. Baba (2021). Environmental and socioeconomic impacts of pipeline transport interdiction in Niger Delta, Nigeria, Heliyon, 7(5) https://doi.org/10.1016/j.heliyon.2021.e06999.

[12] Mahmoodzadeh Z, Wu KY, Lopez Droguett E, Mosleh A (2020). Condition-Based Maintenance with Reinforcement Learning for Dry Gas Pipeline Subject to Internal Corrosion. *Sensors (Basel)*. 2020;20(19):5708. Published 2020 Oct 7. doi:10.3390/s20195708

[13] Y. Guo, B. Zhang, S. Yu, and W. Kai (2020). "Research on an advanced intelligence implementation system for engineering process in industrial field under big data," *Expert Syst. Appl.*, 161.

[14] Salgado, W. L., Dam, R. S. F., Barbosa, C. M., da Silva, A. X., & Salgado, C. M. (2020). Monitoring system of oil by-products interface in pipelines using the gamma radiation attenuation. *Applied Radiation and Isotopes*, *160* (March). https://doi.org/10.1016/j.apradiso.2020.109125

[15] Zainab Almheiri, Mohamed Meguid, Tarek Zayed, Failure modeling of water distribution pipelines using meta-learning algorithms, *Water Research*, 205, 2021, 117680, https://doi.org/10.1016/j.watres.2021.117680.

[16] Doshmanziari, R., Khaloozadeh, H., & Nikoofard, A. (2020). Gas pipeline leakage detection based on sensor fusion under model-based fault detection framework. *Journal of Petroleum Science and Engineering*, 184(July 2019), 106581. https://doi.org/10.1016/j.petrol.2019.106581

[17] Adegbija, T., Lysecky, R., & Kumar, V. V. (2019). Right-provisioned IOT edge computing: An overview. *Proceedings of the ACM Great Lakes Symposium on VLSI, GLSVLSI*, 531–536. https://doi.org/10.1145/3299874.3319338

[18] Adegboye, M. A., Fung, W. K., & Karnik, A. (2019). Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches. *Sensors (Switzerland)*, *19*(11), 1–32. https://doi.org/10.3390/s19112548

[19] Arifin, B. M. S., Li, Z., Shah, S. L., Meyer, G. A., & Colin, A. (2018). A novel data-driven leak detection and

localization algorithm using the Kantorovich distance. *Computers and Chemical Engineering*, *108*, 300–313. https://doi.org/10.1016/j.compchemeng.2017.09.022

[20] P. S. Rajulapati, N. Nukavarapu and S. Durbha, "Deep Learning-based Critical Infrastructure Simulation Model for Disaster Monitoring," *2020 International Conference on Data Mining Workshops (ICDMW)*, 2020, pp. 828-835, doi: 10.1109/ICDMW51313.2020.00136.

[21] S. Zhang *et al.*, "FIGCPS: Effective Failure-inducing Input Generation for Cyber-Physical Systems with Deep Reinforcement Learning," *2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 2021, pp. 555-567, doi: 10.1109/ASE51524.2021.9678832.

[22] D. Fährmann, N. Jorek, N. Damer, F. Kirchbuchner and A. Kuijper, "Double Deep Q-Learning With Prioritized Experience Replay for Anomaly Detection in Smart Environments," in *IEEE Access*, vol. 10, pp. 60836-60848, 2022, doi: 10.1109/ACCESS.2022.3179720.

[23] G. Srikanth, N. Nukavarapu and S. Durbha, "Deep Reinforcement Learning Interdependent Healthcare Critical Infrastructure Simulation model for Dynamically Varying COVID-19 scenario - A case study of a Metro City," *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 8499-8502, doi: 10.1109/IGARSS47720.2021.9554875.

[24] Zakir Ahmad Sheikh, Yashwant Singh, Pradeep Kumar Singh, Kayhan Zrar Ghafoor (2022). Intelligent and secure framework for critical infrastructure (CPS): Current trends, challenges, and future scope, Computer Communications, https://doi.org/10.1016/j.comcom.2022.07.007.

[25] [30] S. K. Prabhakar and S. -W. Lee, "SASDL and RBATQ: Sparse Autoencoder With Swarm Based Deep Learning and Reinforcement Based Q-Learning for EEG Classification," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 3, pp. 58-68, 2022, doi: 10.1109/OJEMB.2022.3161837.

[26] B. Jung, J. Kim and S. Pack, "Deep Reinforcement Learning-based Context-Aware Redundancy Mitigation for Vehicular Collective Perception Services," 2022 International Conference on Information Networking (ICOIN), 2022, pp. 276-279, doi: 10.1109/ICOIN53446.2022.9687254.

[27] [33] S. Bhandari, N. Ranjan, Y. -C. Kim and H. Kim, "Deep Reinforcement Learning for Dynamic Spectrum Access in the Multi-Channel Wireless Local Area Networks," 2022 International Conference on Electronics, Information, and Communication (ICEIC), 2022, pp. 1-4, doi: 10.1109/ICEIC54506.2022.9748733.

[28] Mahmoodzadeh Z, Wu KY, Lopez Droguett E, Mosleh A. Condition-Based Maintenance with Reinforcement Learning for Dry Gas Pipeline Subject to Internal Corrosion. Sensors (Basel). 2020 Oct 7; 20(19):5708. doi: 10.3390/s20195708. PMID: 33036494; PMCID: PMC7582701.

[29] W. L. Keng and L. Graesser, (2019). "Foundations of Deep Reinforcement Learning.", Addison Wesley Professional; 1st edition, 26 – 27, 173 - 176.

[30] Song, S., Kidziński, Ł., Peng, X.B. *et al.* Deep reinforcement learning for modeling human locomotion control in neuromechanical simulation. *J NeuroEngineering Rehabil* 18, 126 (2021). https://doi.org/10.1186/s12984-021-00919-y

[31] Ye, Y.; Qiu, D.; Wang, H.; Tang, Y.; Strbac, G. Real-Time Autonomous Residential Demand Response Management Based on Twin Delayed Deep Deterministic Policy Gradient Learning. *Energies* 2021, *14*, 531. https://doi.org/10.3390/en14030531

[32] Richard S. Sutton and Andrew G. Barto (2018). Reinforcement Learning:An Introduction, second edition, Westchester Publishing Services, 436 – 441.