Utilizing Fractional Order Epidemiological Model to Understand High and Moderate Toxicity Spread on Social Media Platforms*

Emmanuel Addai[0000-0003-2159-6299], Niloofar Yousefi[0009-0005-0071-733X], and Nitin Agarwal[0000-0002-5612-4753]

COSMOS Research Center, University of Arkansas – Little Rock, Arkansas, USA
{eaddai1, nyousefi, nxagarwal}@ualr.edu

Abstract. The COVID-19 pandemic has increased social media usage significantly, highlighting its critical role in public statements, information dissemination, news propagation. In this study, we construct and evaluate a fractional-order toxicity contagion model with quarantine intervention in Twitter. The model incorporates different infected groups that account for toxicity intensity and its development, that is, moderate and high infected users, and it is used to investigate the influence of each user in the overall spread of toxic content. We have evaluated the post-free toxic equilibrium point, the reproduction number ($\mathcal{R}_0$), the existence-uniqueness solution, and the stability point. The model, which fits well to (#F*covid) hashtags data, is qualitatively analyzed to evaluate the impacts of different schemes for control strategies. Our findings reveal that the conventional model, which does not differentiate between infected groups, overestimates or underestimates the rate of change in the number of infectious users, resulting in a greater error rate. From analysis, implementing quarantine measures on social media platforms can bring long-term benefits with low risk, affirming platform safety. By quarantining moderate and high toxic active users, the resulting error rates were impressively low, measured at 0.0011 and 0.0012 for the respective groups of infected users. This study will assist network providers in identifying such users, thereby reducing toxic conversations.

Keywords: toxicity spread · SEIQR epidemiological model · social media · fractional derivative.

1 Introduction

In recent years, online social media have increasingly spread news, election results, and global outbreaks across vast networks, fostering information diffusion

* This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920), U.S. Office of the Under Secretary of Defense for Research and Engineering (FA9550-22-1-0332), U.S. Army Research Office (W911NF-23-1-0011, W911NF-24-1-0078), U.S. Office of Naval Research (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. Air Force Research Laboratory, U.S. Defense Advanced Research Projects Agency, Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

and social connections [1]. The spread of negative content on social media platforms has become a significant issue in today's digital age. With the increasing trend of disseminating information through online social networks, understanding the propagation of toxic content is crucial to effectively mitigate its harmful impacts. Because social media platforms have a significant impact on society [2], it has become increasingly important to gain a deeper understanding of their dynamics. For instance, to investigate the toxicity propagation online, to analyze personal attacks on social media, the authors in [3] employed machine learning methods. Alongside these techniques, several epidemiological models have been developed to analyze online toxicity propagation [4].

Recently, mathematical modeling has become crucial in understanding toxic content and information diffusion on social media, predicting how these processes unfold in a network by learning from past diffusion patterns [5]. Modeling how information or toxic content spreads is critical in stopping the spread of toxic content. Fractional-order epidemiological models have garnered considerable attention in studying disease spread. These models present a unique viewpoint by integrating fractional derivatives to capture memory effects and hereditary characteristics of systems, aspects often disregarded in the traditional models. The utilization of fractional calculus enables a more precise alignment with data and affords additional adaptability in modeling intricate systems [1].

In this work, we construct and evaluate a fractional-order toxicity model with quarantine intervention. The model incorporates different infected groups, that is, moderate and high infected users in the SEI_mI_hQR (Susceptible-Exposed-Moderate Infected-High Infected-Quarantined-Recovered) model, and it is used to investigate the influence of/on each user in the overall spread of toxic content. By dividing the Infected state into moderate and high toxicity levels within the model, we can capture the varying degrees of influence users may apply in propagating toxicity on social media platforms. High-toxic individuals might exhibit behaviors that amplify the spread of toxic content at a faster rate compared to those with moderate toxicity levels. This differentiation allows for a more accurate representation of the dynamics of toxicity propagation. We expect that the fractional-order model memory function will be able to more accurately predict both the dynamics of online behavior in the future and the dynamics of toxicity in the past. This work answers the following research questions: **RQ1:** How can a fractional-order toxicity model with quarantine intervention be constructed and evaluated to understand the dynamics of toxic content spread on social media? **RQ2:** How can the index of memory and the quarantine rate be used as preventive measures against the spread of toxic content on social media networks? **RQ3:** What are the effects of different levels of infection (moderate and high) on the overall spread of toxic content on social media platforms?

2 Related Work

The surge of toxicity on social media has become a major challenge, prompting research into its dynamics, impacts, and solutions. Studies have used machine learning to detect toxic content like hate speech [6], examined the impact of toxicity on Reddit discussions [7], [8]. Additionally, research has explored the effects of toxicity on user engagement and retention in online communities [9].

2.1 Epidemiological modeling with Quarantine State

Epidemiological models for online toxicity liken the spread of toxic content to infectious diseases, assuming toxic behavior can 'infect' users. Researchers use these models to forecast toxic users' behavior [10]. For instance, authors in [4] compare different epidemiological models to study the propagation of toxicity.

Incorporating a quarantine state into epidemiological models significantly enhances our understanding and management of infectious diseases, especially in light of recent global health crises. The inclusion of quarantine—both voluntary and enforced—reflects a critical aspect of disease control strategies. This can significantly alter the disease dynamics, potentially leading to a slower spread of the disease, lower peak prevalence, and ultimately fewer cases. This control strategy has been used to study many diseases such as EBOLA [11] and COVID-19 [12]. It has also been adopted in online social networks; for instance, the authors in [13] investigated effective isolation-based strategies for controlling information spread in social networks. In [14], the authors reviewed models, methods and applications where quarantine control strategy were utilized.

2.2 Fractional Calculus

Fractional-order derivatives are a mathematical tool used to analyze phenomena exhibiting non-linear and complex dynamics, such as the spread of toxicity on social media. Unlike classical integer-order derivatives that describe rates of change in constant time intervals, fractional derivatives provide a more nuanced representation by accounting for memory and hereditary properties of processes [15]. This means they can model the way past interactions and behaviors influence the current spread of toxic content.

Fractional-order derivatives enable models that more accurately capture the intricacies of how toxic behaviors evolve over time on social media. These models can consider the long-range dependencies and varying intensities of interactions among users, which are characteristic of social media dynamics. Fractional derivatives enable the creation of sophisticated models to better capture the complexities of toxicity spread, enhancing strategies for monitoring, predicting, and mitigating harmful content on social platforms.

3 Methodology

This section provides an overview of data collection and the methodology used in this paper.

3.1 Fractional Model Formulation (Caputo Derivative)

Previously, $SEIQR$ (Susceptible-Exposed-Infected-Quarantined-Recovered) was used to study an epidemiological model to contain the spread of toxicity using memory-index [16]. In this work, we propose the SEI_mI_hQR model in the Caputo fractional operator sense to study toxicity spread on social media, which is the extension of the SEIQR model. The proposed SEI_mI_hQR model to study toxicity provides powerful tools for understanding and predicting the spread of toxic contents, especially in scenarios where quarantine is the key strategy for network providers' control. We make several assumptions for the SEI_mI_hQR model: the total number of users is not constant, considering recruitment rate (Π) and autonomous exit rate (μ); the users who spread toxic contents are moderate (I_m) and high infected users (I_h); moderate infections (I_m) can become high infections (I_h) and vice versa at rates κ_m and κ_h, respectively; both moderate and high infections can autonomously recover at rates ϕ_m and ϕ_h, respectively; and recovered users become susceptible again at the rate η.

Therefore, we propose Caputo fractional-order differential equations. For the entire population, we define the quantity N:

$$N(t) = S(t) + E(t) + I_m(t) + I_h(t) + Q(t) + R(t). \tag{1}$$

Hence, we have

Our propose Caputo fractional-order derivative is given as;

$$\begin{cases} {}^{C}D_t^{\alpha}S(t) = \Pi - (\lambda_{mh} + \mu)S + \eta R, \\ {}^{C}D_t^{\alpha}E(t) = \lambda_{mh}S - (\mu + \psi_m + \psi_h)E, \\ {}^{C}D_t^{\alpha}I_m(t) = \psi_m E + \kappa_h I_h - (\mu + \phi_m + \theta_m + \kappa_m)I_m, \\ {}^{C}D_t^{\alpha}I_h(t) = \psi_h E + \kappa_m I_m - (\mu + \phi_h + \theta_h + \kappa_h)I_h, \\ {}^{C}D_t^{\alpha}Q(t) = \theta_m I_m + \theta_h I_h - (\gamma + \mu)Q, \\ {}^{C}D_t^{\alpha}R(t) = \gamma Q + \phi_m I_m + \phi_h I_h - (\mu + \eta)R, \end{cases}$$

where $(\lambda_{mh}) = \frac{\beta(I_m+I_h)}{N}$, $t \in [0, \mathcal{T}]$, $\mathcal{T} \in \mathbb{R}$, and ${}_0^C D_t^{\alpha}$ denotes the Caputo fractional derivative of order, where the memory index is denoted as α, $0 < \alpha \leq 1$, given the initial conditions:

$$S(0) = S\_0 \geq 0, \quad E(0) = E\_0 \geq 0, \quad I_m(0) = I\_m0 \geq 0,$$
$$I_h(0) = I\_h0 \geq 0, \quad Q(0) = Q\_0 \geq 0, \quad R(0) = R\_0 \geq 0.$$

The flow diagram of the model is presented in Fig. 1, while the description of the parameters is presented in Table 1. This table also includes the values that will be used later in our numerical simulations.

It is commonly known that Caputo's derivative offers greater dependability and adaptability in analytical applications. Due to initial condition properties, which are more physically interpretable for most problems, many researchers consider the Caputo operator over other derivatives.

Table 1: Explanation of the model parameters.

Parameter	value	Explanation
Π	100	recruitment rate of human
β	0.0006	effective contact rate
ψ_m	0.009	the rate at which exposed become moderate infected
ψ_h	0.006	the rate at which exposed become high infected
θ_m	0.0067	the rate at which I_m transfer to quarantine class
θ_h	0.017	the rate at which I_h transfer to quarantine class
η	0.04	the rate at which R transfer S
γ	0.002	the rate at which Q transfer to recovery
ϕ	0.1	the rate at which I transfer to recovery
μ	0.1	the rate at which people exit autonomously
κ_m	0.09	the rate at which I_m transfer to I_h
κ_h	0.01	the rate at which I_h transfer to I_m

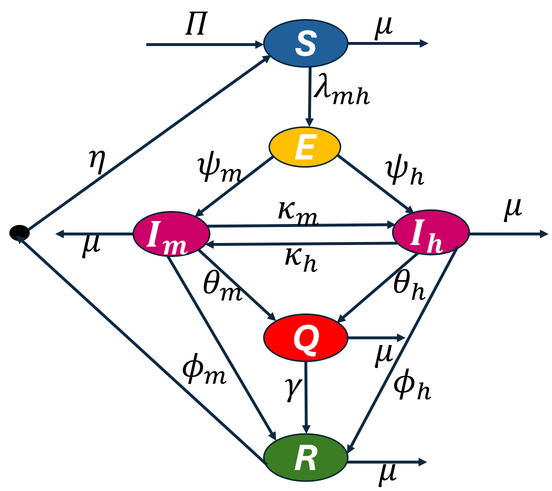


Fig. 1: Transfer diagram for the toxicity spread.

3.2 Data Collection

The Twitter Academic API and the hashtag #Fcovid were used to collect 6,766 COVID-19-related tweets from February 1st, 2020, to December 31st, 2020. A weighted average toxicity of 0.91 was used as a threshold: posts scoring below 0.91 (2,082 posts) were considered moderate, while those scoring above 0.91 (4,684 posts) were deemed highly toxic. To identify the most active users among the infected, we focus on the retweets of posts shared by individuals within the subset of high and moderate toxic users. A higher number of retweets indicates active engagement with the post.

4 Model Analysis

In this section, we conduct the qualitative analysis of the proposed model.

4.1 Basic Reproduction Number

The basic reproduction number represented as $\mathcal{R}_0$, is described as the average number of secondary infections generated by one infected individual within a completely susceptible population. The primary application of $\mathcal{R}_0$ lies in assessing whether a new infectious disease has the potential to proliferate among a population. For the SEI_mI_hQR model, the basic reproduction number $\mathcal{R}_0$ is

$$\mathcal{R}_0 = \frac{\psi_m \beta \mathcal{A}}{\mu(\mu+\psi_m)(\mu+\phi_m+\theta_m)} + \frac{\psi_h \beta \mathcal{A}}{\mu(\mu+\psi_h)(\mu+\phi_h+\theta_h)}.$$

4.2 Existence-Uniqueness and Stability Analysis

This section examined the existence-uniqueness and stability concept of our proposed model. We study the following Theorems to achieve this goal.

Theorem 1. *The Caputo fractional toxicity spread model has a unique solution under the condition that*

$$\frac{\mathcal{T}^\alpha}{\Gamma(\alpha)}\mathcal{L}_i < 1,\ \ i = 1, 2, ..., 5$$

when $t \in [0, \mathcal{T}]$ *and* $\mathcal{L}_i$ *satisfy the Lipschitz condition.*

Proof. From Lemma 0.1 and applying the proof from Theorem 7 of [17], our proposed toxicity spread model exists and has a unique solution.

Theorem 2. *The Caputo fractional toxicity spread model is Hyer-Ulam stable, if there exists*

$$\frac{\mathcal{T}^\alpha}{\Gamma(\alpha+1)}\mathcal{L}_i < 1,\ \ i = 1, 2, ..., 5$$

Proof. From Lemma 0.1 and applying the proof from Theorem 6.2 of [15], our proposed toxicity spread model is Hyers-Ullam stable.

5 Model Parameterization and Data Fitting Analysis

The least-squares technique is the most commonly used method to estimate parameter values. It is defined as:

$$E_{-\mathrm{rel}} = \frac{\|I_{est}(t_i) - I_{data}(t_i)\|_2}{\|I_{data}(t_i)\|_2}. \tag{2}$$

The relative error 2 is applied to compute the model's error. We rank users according to the average toxicity score of posts and the number of retweets they received. Tables 2 and 3 show users with higher retweet counts and moderate to high average toxicity scores in their posts. These tables indicate the model's error by systematically removing each user one by one. We removed five users based on their retweet counts and average toxicity scores for both moderate and high toxicity groups. The error remains consistent after the removal of 5th user. We aim to minimize user removal while achieving optimal results.

Table 2: Examining the error impact of removing posts from moderate most active and most toxic users.

User	$Error_r$	Average toxicity score	Number of Retweet Received
1^{st}	0.0031	0.883	134
2^{nd}	0.0027	0.862	75
3^{rd}	0.0021	0.881	27
4^{th}	0.0011	0.873	10
5^{th}	0.0011	0.858	8

Table 3: Examining the error impact of removing posts from high most active and most toxic users.

User	$Error_r$	Average toxicity score	Number of Retweet Received
1^{st}	0.011	0.998	1489
2^{nd}	0.009	0.939	227
3^{rd}	0.0073	0.956	67
4^{th}	0.0012	0.997	67
5^{th}	0.0012	0.997	44

6 Numerical Simulation and Discussion

This section presents numerical results showing system trajectories with varying input parameters. We examine different scenarios to understand the system's behavior under diverse conditions. Our analysis of these numerical data is focused on highlighting the system's adaptability and vulnerability to various input changes. For our numerical analysis, it is important to acknowledge that certain assumptions have been made regarding the initial conditions of the state variables ($S = 12000; E = 11000;$ $I_m = 3000; I_h = 3000; Q = 7000; R = 2000$) and system parameters (refer to Table 1).

Simulations were conducted to visualize the SEI_mI_hQR model under varying parameters. By plotting the trajectories of toxicity spread across all the

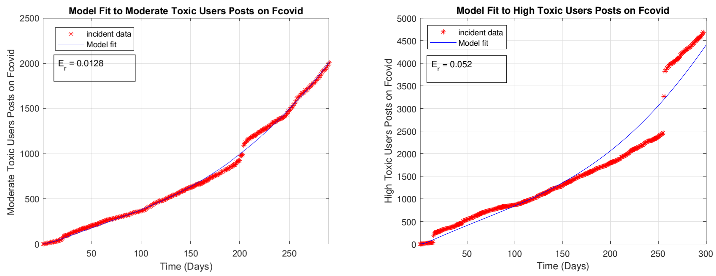


Fig. 2: Fitted SEI_mI_hQR model for moderate/high toxic users posts.

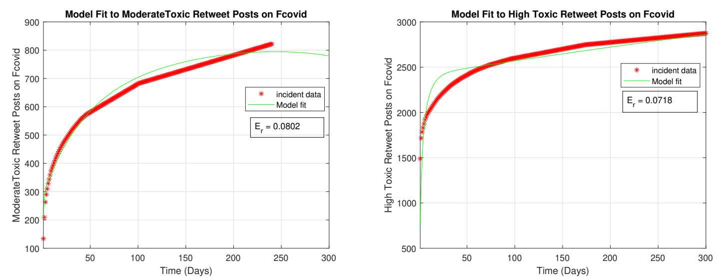


Fig. 3: Fitted SEI_mI_hQR model for moderate/high toxic users with high retweets received.

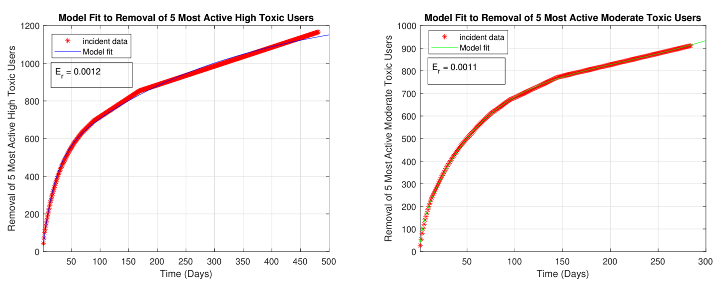
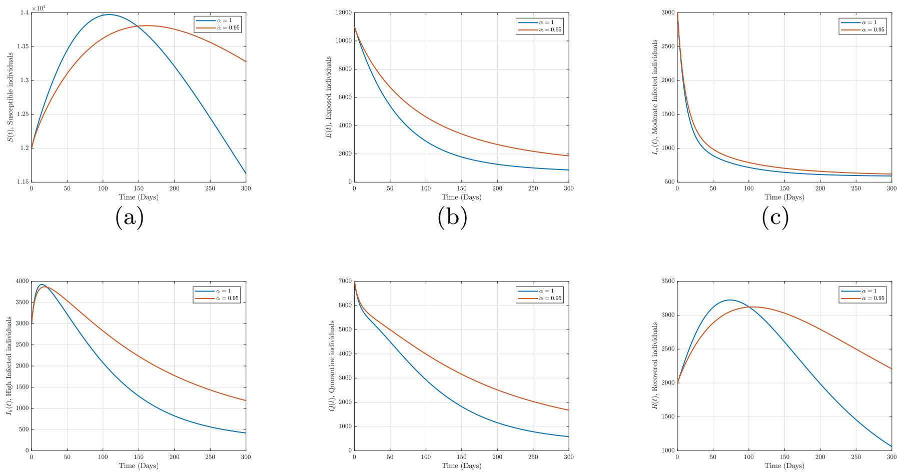


Fig. 4: Fitted SEI_mI_hQR model for the top 5 high/moderate most active and toxic users transferred to quarantine.

compartments over time, we can understand the progression of toxic posts comprehensively. First, from Fig. 2 - 4, the data is observed to depict the toxicity spread. The data is compared with the proposed model to check the accuracy of the proposed model. Fig. 2 shows cumulative moderate toxic posts against the proposed model's moderate infected (I_m) with an observed error of 0.0128. similarly, for high toxic posts with an observed error of 0.052. Fig. 3, provides insights into how the model fits the retweet activity of moderate and high toxic users, the lower error of 0.0802 and 0.0718, respectively.



(a) (b) (c)

(d) (e) (f)

Fig. 5: Numerical dynamics on SEI_mI_hQR model at different fractional order α.

In Fig. 4, our proposed model, which involves transferring users with high and moderate average toxicity, the top five users were removed from moderate and high infected and transferred to quarantine. The resulting errors were 0.0011 and 0.0012, respectively. This result provides evidence that our proposed model, involving grouping the infected into moderate and high and then transferring some of the users to quarantine, yields superior outcomes in error reduction.

The plots in Fig. 5 present simulations of the memory effect using the Adams –Bashforth–Moulton Predictor–Corrector method. Fig. 5a and 5b shows the impact of fractional parameters on the Susceptible and Exposed classes. This behavior indicates a large influx of new users and high exposure due to the platform's fractional operator and toxic nature. In Fig. 5c and Fig. 5d, we illustrate the moderate and high infected classes, respectively, and begin to increase them by lowering the fractional values. Fig. 5e shows increased dynamics due to the number of infected individuals increasing, as depicted in Fig. 5c and 5d, and hence a high number of users to quarantine. In Fig. 5f, we observed that the model captured high recovery due to interactions among the network's toxic users and the introduction of quarantine as a control intervention.

7 Conclusion

In this study, we investigated toxicity spread on social media platform using the SEI_mI_hQR epidemiological model, taking social network service provider intervention into account and quarantining users, meanwhile, considering the index of memory nature of parameters of the model, which makes it more appropriate to the real situation. The post-free toxic equilibrium point exists for the model. The reproduction number ($\mathcal{R}_0$) is determined using the next-generation matrix method, and the existence-uniqueness, as well as stability, are also analyzed using fractional operator techniques. Our findings from this study reveal that the conventional model, which does not differentiate between infected groups, either overestimates or underestimates the rate at which individuals move from the Exposed compartment to the Infected and then to the Quarantine during the rising or falling phases of the number of infectious users, respectively.

Implementing quarantine measures by network service providers can be highly beneficial, showing a low level of risk and affirming platform safety. Quarantining toxic and active users resulted in impressively low error rates of 0.0011 and 0.0012 for the moderate and high toxic users, respectively. Furthermore, the index of memory and the quarantine rate can be used as preventive measures for the toxicity spread on social network. Future research will use different data and methods to improve this model and explore additional control measures.

References

1. Dai, L., Liu, X., Chen, Y.: Global dynamics of a fractional-order SIS epidemic model with media coverage. In: Nonlinear Dyn. vol. 11, pp. 19513–19526, (2023).
2. Shajari, S., Alassad, M., Agarwal, N.: Characterizing Suspicious Commenter Behaviors. In: International Conference on Advances in Social Networks Analysis and Mining, (2023).
3. Wulczyn, E., Nithum, T., Dixon, L.: Ex machina: Personal attacks seen at scale. In: The 26th International Conference on World Wide Web (WWW 2017), pp. 1391–1399, (2017). doi=10.1145/3038912.3052591
4. N. Yousefi, N., Agarwal, N.: Study the Influence of Toxicity Intensity on Its Propagation Using Epidemiological Models. In: the proceeding of 30th Americas Conference on Information Systems, (2024).
5. Bettencourt, L. M., Cintrn-Arias, A., Kaiser, D. I., Castillo-Chvez, C.: The power of a good idea: quatitative modeling of the spread of ideas from epidemiological models. In: Physica A vol. 364, pp. 513–536, (2006).
6. Yousefi, N., Noor, N. B., Spann, B., Agarwal, N.: Towards developing a measure to assess contagiousness of toxic tweets. In: Workshop Proceedings of the 17th International AAAI Conference on Web and Social Media, TrueHealth 2023: Workshop on Combating 13 Health Misinformation for Social Wellbeing, 5-8 June, 2023, Limassol, Cyprus, p. 43. (https://doi.org/10.36190/2023.43)
7. Yousefi, N., Noor, N. B., Spann, B., Agarwal, N.: Examining Toxicity's Impact on Reddit Conversations. In: Complex Networks & Their Applications XII, H. Cherifi, L. M. Rocha, C. Cherifi, and M. Donduran (eds.), Cham: Springer Nature Switzerland, pp. 401–411, (2023). (https://doi.org/10.1007/978-3-031-53503-1_33)
8. Falade, T. C. C., Yousefi, N., Agarwal, N.: Toxicity Prediction in Reddit. In: the Proceeding of 30th Americas Conference on Information Systems, (2024).
9. Duggan, M.: Online Harassment. Accessed: 2023-12-24, (2023).
10. Yousefi, N., Agarwal, N., Addai, E.: Developing Epidemiological Models with Differentiated Infected Intensity. In: 17th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation, In press, (2024).
11. Ngwa, G. A., Teboh-Ewungkem, M. I.: A Mathematical Model with Quarantine States for the Dynamics of Ebola Virus Disease in Human Populations. In: Comput. and Maths Methods in Medicine, 9352725, (2016).

12. Alsheri, A. S., Alraeza, A. A., Afia, M. R.: Mathematical modeling of the effect of quarantine rate on controlling the infection of COVID-19 in the population of Saudi Arabia. In: Alexandria Engineering Journal, 61(9), pp. 6843–6850, (2022).
13. Li, T., Wang, S., Li, B.: Research on Suppression Strategy of Social Network Information Based on Effective Isolation. In: Procedia Computer Science, 131, pp. 131–138, (2018).
14. Chen, B. L., Jiang, W. X., Chen, Y. X., Chen, L., Wang, R., Han, S., Lin, J. H., Zhang, Y. C.: Influence blocking maximization on networks: Models, methods and applications. In: Physics Reports, 976, pp. 1–54, (2022).
15. Zhang, N., Addai, E., Zhang, L., et al.: Fractional modeling and numerical simulation for unfolding Marburg-monkeypox virus co-infection transmission. In: Fractals. 31 (2350086), (2023). https://doi.org/10.1142/S0218348X2350086X
16. Addai, E., Yousefi, N., Agarwal, N.: SEIQR: An Epidemiological Model to Contain the Spread of Toxicity using Memory-Index. In: Fifth International Workshop on Cyber Social Threats, International Conference on Web and Social Media, In press, (2024).
17. Addai, E., Ngungu, M., Omoloye, M. A., Marinda, E.: Modelling the impact of vaccination and environ-mental transmission on the dynamics of monkeypox virus under Caputo operator. In: AIMS MBE, (2023).