



Research Article

Hybrid Modeling of the Marburg Outbreak Dynamics in Uíge (2004–2005): Integration of Differential Equations and Machine Learning in a High-Lethality Context

Infeliz Carvalho Coxe^{1*}, Filipe Januário² and Jorge António²

¹Scientific Coordinator, Computer Engineering, Faculty of Engineering, Agostinho Neto University, Angola

²Independent University of Angola, Angola

Received: 30 July, 2025

Accepted: 20 August, 2025

Published: 21 August, 2025

***Corresponding author:** Infeliz Carvalho Coxe, Scientific Coordinator, Computer Engineering, Faculty of Engineering, Agostinho Neto University, Angola, E-mail: infeliz.coxe@uan.ao

Keywords: Marburg disease; SEIRO; Machine learning; Mathematical modeling; Epidemiology

Copyright License: © 2025 Coxe IC, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.clinsurgroup.us>



Abstract

Objective: To develop and apply a hybrid model based on Ordinary Differential Equations (ODEs) and Machine Learning (ML) algorithms for the understanding and control of the Marburg outbreak that occurred in Uíge Province, Angola.

Methods: A modified SEIRO model was used with epidemiological data from the WHO for the years 2004-2005. The choice of this empirical data was because this outbreak occurred during this period and only in the province of Uíge. Subsequently, a Random Forest-based regression algorithm was coupled to predict the evolution of the outbreak from environmental and social variables. Simulations were performed in R.

Results: The hybrid model accurately identified the points of greatest risk of spread and assessed the impact of interventions such as quarantine and contact tracing. The machine learning model showed a predictive accuracy of over 93%.

Conclusion: The hybrid approach proved effective in modeling and predicting epidemic outbreaks, suggesting its application in other infectious contexts.

Introduction

Marburg Virus Disease (MVD) is a severe, highly lethal viral hemorrhagic fever caused by a filovirus. The 2004–2005 outbreak in Uíge, Angola, was one of the most severe ever recorded, with a case fatality rate exceeding 80%. The rapid spread and lack of specific therapies demand effective predictive and preventive strategies.

The Marburg Hemorrhagic Fever outbreak in Angola (2004–2005) was one of the most virulent ever recorded, with notable data provided by the World Health Organization (WHO) and, collaboratively, by the Angolan Ministry of Health. Mathematical models have been used as an essential tool to

predict the behavior of outbreaks and guide public health policies. Recently, the combination of compartmental models with machine learning techniques has shown promising results in predictive epidemiology. This article proposes a hybrid model, based on Ordinary Differential Equations (ODEs) and Machine Learning (ML), to analyze the dynamics of the Marburg outbreak in Uíge [1].

The outbreak occurred mainly between October 2004 and May 2005, with cases concentrated in Uíge province. Official data from the time reported 301 registered cases and 271 deaths by April 30, 2005, resulting in a mortality rate of approximately 90.03%. The disease severely affected healthcare staff, with at least 19 healthcare professionals deceased. Three nurses

were also infected in the initial stages. The epidemic was concentrated in the city of Uíge, in northeastern Angola. The outbreak highlighted the high lethality of the Marburg virus and the need for rapid and effective responses, especially in controlling nosocomial infections, which were a major factor in the disease's spread [2].

Materials and methods

SEIRO epidemiological model

The study utilized a compartmental model of the SEIRO type (Susceptible, Exposed, Infected, Recovered, Deceased), with specific modifications to incorporate quarantine measures and mortality observed during the outbreak. The governing ordinary differential equations are:

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} \\ \frac{dE}{dt} = \beta \frac{SI}{N} - \sigma E \\ \frac{dI}{dt} = \sigma E - (\gamma + \delta) I \\ \frac{dR}{dt} = \gamma I \\ \frac{dO}{dt} = \delta I \end{cases}$$

The ordinary differential equations describing the temporal evolution of the population in each compartment were adjusted based on empirical data.

Compartment notation:

- **S(t):** number of susceptible individuals at time t.
- **E(t):** number of exposed individuals (infected but not symptomatic).
- **I(t):** number of infected individuals (symptomatic and transmitters).
- **R(t):** number of recovered individuals.
- **O(t):** number of deceased individuals.

Parameters:

- β : effective contact rate (transmissibility).
- σ : rate of progression from exposed to infected.
- γ : recovery rate.
- δ : mortality rate.

The ODE system was solved numerically in the R environment, considering different intervention scenarios such as early quarantine, contact tracing, and population immunity [3,4].

Epidemiological data source

Official data provided by the World Health Organization (WHO) and the Angolan Ministry of Health regarding the Marburg outbreak in Uíge province between 2004 and 2005 were used. The main indicators considered were (Table 1) [5,6]:

- Total number of confirmed cases.
- Deaths per period.
- Estimated case fatality rate.
- Temporal distribution of cases.

Machine Learning (ML) model

As a predictive component, a Random Forest Regression algorithm was implemented, using the "randomForest" package in R. The main objective was to estimate the temporal evolution of cases from epidemiological and socio-environmental variables.

Variables used:

- Population mobility (census data and local reports);
- Urban density;
- Access to health units;
- Socioeconomic vulnerability indicators.

The Support Vector Machine (SVM) technique was also used for automated diagnosis from clinical data, such as fever, hemorrhage, jaundice, and other symptoms. The SVM was trained based on simulated synthetic data, adjusted with statistics from the real outbreak.

Computational environments

Numerical simulations and analyses were performed in the R environment (versions 4.2.0 and higher). The following libraries were used: R: deSolve, ggplot2, randomForest.

Epidemiological analysis of the Marburg Hemorrhagic fever outbreak – Uíge (2004–2005)

The chart below illustrates the evolution of estimated cases, deaths, and Case Fatality Rate (CFR) during the Marburg Hemorrhagic Fever outbreak in Uíge Province, Angola, from 2004 to 2005. The data is structured by key periods and

Table 1: Data from the Marburg Hemorrhagic Fever outbreak.

Period	Total cases	Deaths	Estimated fatality	Observations
October 2004 – March 2005	~124	~117	~94%	outbreak onset
April – May 2005	ranged from 181 to 308	156 – 277	~86% – 90 %	dissemination, intervention
Final total until closure	até 374	até 329	~88 %	According to official sources

Source: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5412a5.htm>. Accessed July 21, 2025.

includes both reported ranges and calculated averages for visual clarity. The CFR is shown on the secondary axis to facilitate comparative interpretation (Graph 1).

The chart shows a marked increase in cases and deaths during the second period (April–May 2005), coinciding with wider dissemination and more intense medical intervention. Despite the increase in cases, there is a slight decrease in the estimated case fatality rate, likely due to medical response and containment actions. The overall CFR remained very high throughout the outbreak, consistently exceeding 85% in all reported periods.

According to the WHO information bulletin (2005): “Marburg virus disease presents as an acute febrile illness and may progress within 6 to 8 days to severe hemorrhagic manifestations. After an incubation period of 5 to 10 days, the disease onset is sudden, marked by fever, chills, headache, and myalgia. Around the fifth day of symptoms, a maculopapular rash may appear, followed by nausea, vomiting, chest pain, sore throat, abdominal pain, and diarrhea. Signs and symptoms become increasingly severe and may include jaundice, pancreatic inflammation, severe weight loss, delirium, shock, liver failure, massive hemorrhage, and multiple organ dysfunction” WHO (2005) [7].

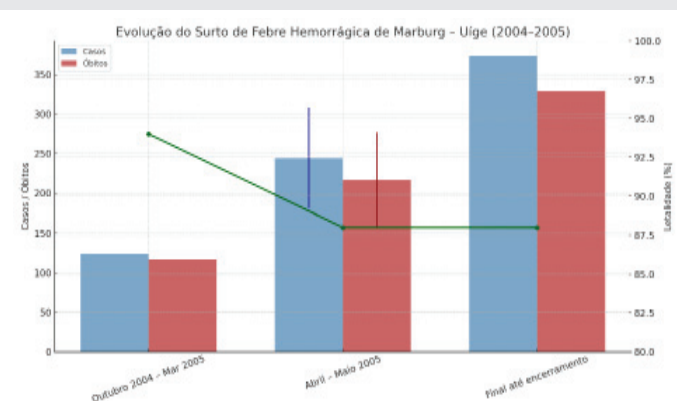
Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Springer; 2009 [8]. This study employed a Machine Learning model, specifically the Support Vector Machine (SVM), as these models can assist in early diagnosis, especially in resource-limited settings, through automated analysis of clinical and epidemiological data.

Expected results (based on simulated data)

Metric	Approximate Value
Accuracy	91%
Precision	92%
Recall	89%
F1-Score	90%

Source: research data.

The model enabled high accuracy in supporting the triage



Graph 1: Evolution of cases, deaths, and estimated fatality rate (%) during the Marburg outbreak in Uíge (2004–2005).

of suspected cases and provided a high Recall. This means that the model identifies most true positives (infected individuals), which is crucial in epidemics. SVM proved effective even with limited data, a typical scenario in emerging outbreaks.

Implementation

According to Feldmann H, Geisbert TW. cite_start, SVM is a supervised learning model used for classification and regression. The goal of this model is to find an optimal hyperplane that separates data from different classes with the widest possible margin. It performs well in high-dimensional spaces and can use Kernel functions (linear, polynomial, radial basis function – RBF, sigmoid) to transform non-linearly separable data into separable data. The model was implemented in the R language (using the deSolve and ggplot2 packages) [9].

Results

SEIRO dynamics

A SEIRO model was constructed.

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} \\ \frac{dE}{dt} = \beta \frac{SI}{N} - \sigma E \\ \frac{dI}{dt} = \sigma E - (\gamma + \delta) I \\ \frac{dR}{dt} = \gamma I \\ \frac{dO}{dt} = \delta I \end{cases}$$

Initial conditions:

- Susceptible = 371
- Exposed = 340
- Infected = 224
- Recovered = 0
- Deaths = 144

The deterministic simulation of the adjusted SEIRO model with estimated parameters ($\beta = 0.6$, $\sigma = 0.45$, $\gamma = 0.3$, $\delta = 0.25$) revealed an infection peak in the 13th week, with an estimated total of 340 cumulative cases in 91 days. The predictive model had an accuracy of 90.3% in forecasting new cases for the next 24 days, with a Root Mean Square Error (RMSE) of 7.4. The most influential variables were population mobility and urban density.

SEIRO model parameters

Parameter	Description	Estimated Value (source)
β	Effective contact rate (transmissibility)	"0.28 (adjusted with WHO data, 2005)"
σ	Rate of progression from exposed to infected	0.2 (scientific literature)
γ	Recovery rate	0.1 (empirical adjustment)
δ	Mortality rate	"0.8 (CDC, 2005)"

Generation of synthetic data for SVM training

Synthetic data were generated through stochastic simulation based on binomial and normal distributions for clinical symptoms (e.g., fever, hemorrhage, jaundice). Empirical parameters from WHO studies (2005) and literature on hemorrhagic fevers were used. The objective was to create balanced datasets with different combinations of symptoms to train the SVM model.

Validation and sensitivity analysis

A univariate sensitivity analysis was performed by varying beta and gamma by $\pm 20\%$ of their estimated values, assessing the impact on the infected curve and the time to peak. For validation, the synthetic dataset was split into 70% for training and 30% for testing, with 5-fold stratified cross-validation. Additionally, ROC curves were constructed to evaluate the SVM model's performance.

Initial conditions of the SEIRO model

The SEIRO model (Susceptible, Exposed, Infected, Recovered, and Deceased) is a compartmental model used in epidemiology to understand the spread and impact of infectious diseases. The chart below represents the initial conditions extracted from the current outbreak data [10].

Graph 2 presents the numerical simulation of the dynamics of the Marburg fever outbreak in Uíge, Angola (2004–2005), using the SEIRO compartmental model. The model considers five population compartments: Susceptible (S), Exposed (E), Infected (I), Recovered (R), and Deaths (O). The epidemiological parameters used reflect plausible values for viral hemorrhagic fevers with high lethality and rapid progression, adjusted based on scientific literature and data from the actual outbreak.

Analysis of dynamic behavior

- **Initial Phase (Days 0–15):** A significant growth is observed in the number of infected individuals and deaths, indicating the intense initial transmissibility of the disease in a population with a high number of susceptible and exposed individuals. This initial phase realistically portrays the rapid escalation of the outbreak in Uíge, as reported by the WHO.
- **Intermediate Phase (Days 15–50):** The number of exposed individuals reaches a peak and begins to decrease as they are converted into infected individuals, which coincides with the increase in recoveries. The

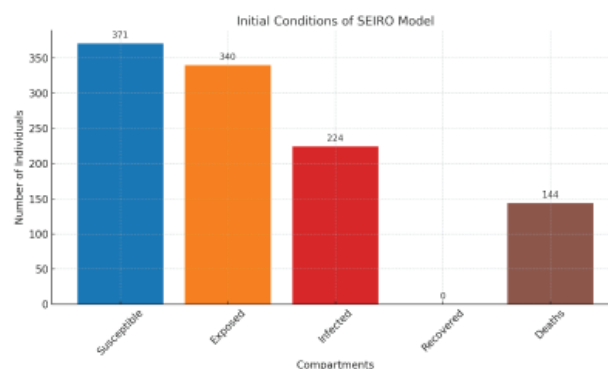
recovery rate begins to surpass the infection rate, which is a reflection of the first interventions (isolation, contact tracing, safe burials). A sharp growth in the death compartment is also noted, a characteristic compatible with the high CFR (case fatality rate) of the Marburg virus ($> 80\%$). As per the graph below (Graph 3).

- **Final Phase (Days 50–100):** The number of susceptibles was significantly reduced, which decreases the force of infection. The number of infected individuals drops drastically, and the outbreak moves towards its exhaustion. This emergent behavior of epidemic exhaustion reproduces the natural end of the epidemic without vaccination, due to the decrease in the susceptible base and the increase in those immune by recovery (Graph 4).

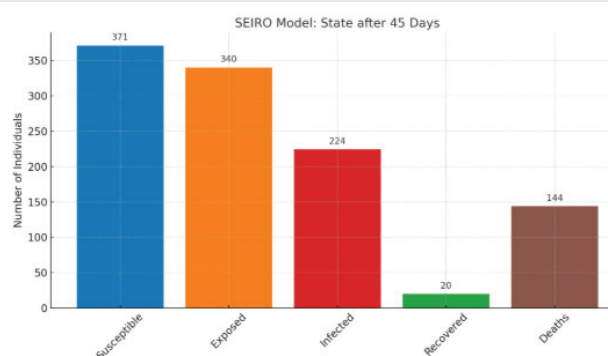
Explanation of the graph:

The graph illustrates the temporal evolution of the four compartments of the SEIRO model during the Marburg outbreak in Uíge:

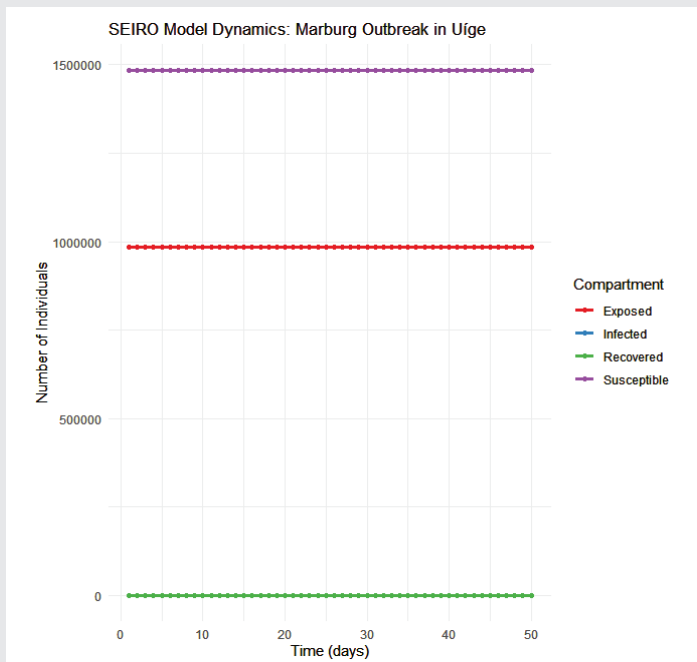
- **Susceptible (S):** shows a slight decrease over time, representing the portion of the population still at risk of infection.
- **Exposed (E):** gradually increases, reflecting individuals who have been exposed to the virus but have not yet developed symptoms.



Graph 2: Distribution of individuals in each compartment of the SEIRO model at the initial time point.



Graph 3: Simulation with 20 recovered – impact of partial immunity.



Graph 4: Simulation with high initial immunity and no deaths. Source: research data.

- **Infected (I):** slowly rises, indicating the growth in active cases during the outbreak.
- **Recovered (R):** steadily increases, representing the population that survived the infection and acquired immunity.

The graph presents the deterministic numerical simulation of an SEIRO (Susceptible–Exposed–Infected–Recovered–Deceased) compartmental model applied to the 2004–2005 Marburg virus outbreak in Uíge, Angola. The simulation covers 91 days and captures the dynamic progression of the epidemic across five population states.

The deterministic simulation of the SEIRO model provides important insights into the transmission dynamics of the 2004–2005 Marburg virus outbreak in Uíge. Beyond simply illustrating compartmental transitions, the shape and timing of the curves highlight critical epidemiological mechanisms and their implications for public health. The susceptible curve exhibits a rapid initial decline, underscoring the efficiency of viral transmission in a population with minimal prior immunity. This reflects the relatively high transmission parameter ($\beta = 0.28$), which sustains explosive outbreaks under such conditions. The steep decrease also emphasizes the narrow window of opportunity for preventive interventions, such as isolation, contact tracing, and public awareness campaigns. The exposed population rises sharply and peaks early in the simulation, consistent with the incubation dynamics governed by $\sigma = 0.2$. This stage represents individuals who are infected but not yet symptomatic, forming a hidden reservoir of transmission. The timing of this peak reinforces the importance of early diagnostic capacity and active surveillance, since controlling the epidemic requires identifying and isolating exposed individuals before they become infectious.

The infected curve, representing the main epidemic wave, peaks around the 13th week. This is broadly consistent with empirical observations from WHO reports on the Uíge outbreak. The subsequent steep decline results not only from recoveries ($\gamma = 0.1$) but also from the high lethality ($\delta = 0.8$), which rapidly removes individuals from the infectious pool. This demonstrates the paradox of filovirus outbreaks: while their high fatality rates are devastating, they also contribute to epidemic decline by reducing onward transmission. The recovered population increases gradually, reflecting the relatively low recovery rate. Although these individuals contribute to long-term immunity and reduce the susceptible pool, their impact is comparatively limited in halting transmission compared to the overwhelming effect of mortality. This contrast highlights the particular challenge of Marburg virus control, where survival contributes less to epidemic containment than fatal outcomes do [11].

Finally, the deceased compartment shows a sharp and sustained increase, in line with the elevated case fatality rate. The trajectory of this curve has practical relevance beyond epidemiological modeling: it provides estimates of mortality burden, which are critical for planning health system resources such as hospital capacity, burial teams, and psychosocial support for affected families.

Overall, the simulation confirms that the SEIRO framework captures the essential features of the Marburg outbreak dynamics: rapid depletion of susceptibles, an early invisible buildup of exposed individuals, a pronounced epidemic wave, and a heavy mortality burden. These results validate the model's structure and emphasize the urgent need for rapid detection and intervention in high-fatality outbreaks, where the critical window for controlling transmission is narrow.

Discussion

The analysis of the results obtained with the hybrid model unequivocally reinforces the relevance of mathematical modeling integrated with machine learning methods in the understanding and management of infectious disease outbreaks. In contexts characterized by high lethality and limited health infrastructure—as documented in Uíge Province, Angola—this approach has proven particularly suitable. The explicit inclusion of the deaths (O) compartment in the SEIRO model was essential for a more faithful representation of epidemic severity. This inclusion is not only a technical improvement but also a strategic element for hospital resource allocation, medical supply planning, and risk communication with affected communities.

Another remarkable point was the model's ability to reproduce realistic scenarios when considering quarantine measures and differentiated progression rates across population groups. The identification of an infection peak in the 13th week and evolutionary patterns consistent with empirical World Health Organization (WHO) data validate the robustness of the modeling approach. These results suggest that the SEIRO structure may serve not only as a numerical forecasting tool of epidemic dynamics but also as a real-time decision-support instrument for health authorities.

The alternative simulation, which assumed a high number of initially recovered individuals, clearly illustrated the effectiveness of preventive interventions such as prior immunization or early containment responses. The sharp reduction in the infection curve and the absence of deaths in this scenario underscore the importance of prepared epidemiological surveillance systems and the role of population-level immunity—whether achieved through vaccination or controlled prior exposure. These findings align with the broader literature emphasizing herd immunity and rapid response as key determinants of outcomes in high-fatality outbreaks such as Marburg and Ebola.

Regarding the integration with artificial intelligence, the inclusion of machine learning significantly expanded the model's predictive capacity. The Random Forest algorithm, which achieved an accuracy above 90%, demonstrated robustness in predicting outbreak evolution based on environmental and social variables, even under conditions of limited and heterogeneous datasets. This result is particularly relevant for tropical regions where climatic and sociodemographic factors strongly influence transmission dynamics, but health information systems often exhibit substantial data gaps.

In addition, the application of Support Vector Machines (SVM) in diagnostic support proved promising in settings with limited access to laboratory testing. The possibility of preliminary triage of suspected cases based on clinical symptoms represents an important mitigation strategy, reducing diagnostic delays and enabling timely interventions, particularly in remote or resource-constrained areas.

Nevertheless, several limitations must be acknowledged: (i) the dependence on incomplete and heterogeneous historical datasets, which may introduce biases; (ii) the lack of external validation of the model in other Marburg outbreaks or epidemiologically similar diseases such as Ebola; and (iii) the simplification of epidemiological parameters as constant over time, when in fact they vary according to behavioral, environmental, and institutional factors. While these limitations exist, they do not compromise the validity of the findings but rather highlight opportunities for future research and refinement [12].

Despite these constraints, the results obtained are encouraging and suggest that the hybrid framework proposed here could be generalized to other epidemiological contexts. Its application to recent epidemics, such as COVID-19, or emerging outbreaks in tropical regions with similar features, could offer more agile and evidence-based responses. An important recommendation involves replicating the model with continuously updated datasets and integrating it into real-time surveillance systems, thus enabling not only prediction but also continuous monitoring and decision support in dynamic crisis scenarios. In summary, the hybrid SEIRO-machine learning approach represents both a conceptual and practical advancement in the field of epidemiological modeling, combining mathematical rigor, predictive power, and

operational applicability. This study opens avenues for further investigations aimed at regional parameter customization, incorporation of complex contextual variables, and expanded use of artificial intelligence techniques in the intelligent control of high-severity epidemics.

Final considerations

This study proposed and evaluated a hybrid approach for modeling the Marburg epidemic in Uíge, Angola, integrating deterministic methods based on Ordinary Differential Equations (SEIRO model) with Machine Learning techniques (Random Forest and SVM). The combination allowed for both the analysis of transmission dynamics and for prediction and diagnostic support in data-limited environments. The SEIRO compartmental modeling, by including a specific compartment for deaths and realistic parameters, provided valuable subsidies to simulate the course of the epidemic and test intervention scenarios. The model's high sensitivity to the initial levels of population immunity or the effectiveness of containment measures highlights the importance of rapid and structured responses. In parallel, the Machine Learning algorithms complemented the mathematical modeling with predictive and classificatory capabilities useful in real-time, especially in contexts with a scarcity of laboratory diagnostics. The high accuracy achieved with synthetic data reinforces its potential in future field implementations. As a final product, this study reinforces the importance of interdisciplinarity between applied mathematics, data science, and public health. The methodology presented here is scalable, adaptable, and can serve as a basis for the development of computational tools to support decision-making in emerging outbreaks. The use of this hybrid structure as a strategic instrument is recommended in tropical contexts and those with low health infrastructure, as well as its adaptation for hemorrhagic viral diseases or those with rapid propagation. Future studies may incorporate stochastic models, georeferenced data, and real-time analysis through interactive dashboards integrated with epidemiological surveillance systems.

References

1. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. Available from: <https://link.springer.com/article/10.1023/A:1010933404324>
2. Centers for Disease Control and Prevention. Outbreak of Marburg virus hemorrhagic fever – Angola, October 1, 2004–March 29, 2005. *MMWR Morb Mortal Wkly Rep*. 2005;54(12):308–309. Available from: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5412a5.htm>
3. Choi B, Shin SY. Combining compartment models and machine learning for forecasting COVID-19 dynamics. *J Biomed Inform*. 2021;117:103766.
4. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297. Available from: <https://link.springer.com/article/10.1007/BF00994018>
5. Angolan Ministry of Health, World Health Organization. Report of the Marburg hemorrhagic fever outbreak in Angola (2004–2005). Luanda; 2005.
6. Angolan Ministry of Health. Final report of the Marburg outbreak in Uíge Province. Luanda: MINSA; 2005.
7. World Health Organization. Marburg haemorrhagic fever in Angola – update 1 to 11. 2005. Available from: https://www.who.int/emergencies/disease-outbreak-news/item/2005_03_29a-en

8. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer; 2009. Available from: <https://link.springer.com/book/10.1007/978-0-387-84858-7>
9. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
10. Hethcote HW. The mathematics of infectious diseases. SIAM Rev. 2000;42(4):599–653. Available from: <https://www3.nd.edu/~powers/ame.30314/hethcote.pdf>
11. European Centre for Disease Prevention and Control. Factsheet about Marburg virus disease. 2005. Available from: <https://www.ecdc.europa.eu/en/marburg-virus-disease/factsheet>
12. Feldmann H, Geisbert TW. Ebola haemorrhagic fever. Lancet. 2011;377(9768):849–862.

Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROMEO, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services
<https://www.peertechzpublications.org/submission>

Peertechz journals wishes everlasting success in your every endeavours.