IAENG International Journal of Computer Science

Epidemiological Insights and Forecasting of HIV/AIDS Mortality: A Case Study of South Africa and Myanmar Using Machine Learning Approaches

Wenbin Yang, Xin Chang

Abstract—**HIV/AIDS is one of the deadliest infectious diseases in the world, causing nearly one million deaths annually. In some countries, HIV/AIDS is the leading cause of death, especially in regions like South Africa. In the Asian region, Myanmar is also a high-prevalence area for HIV/AIDS-related deaths. This study utilized data on HIV/AIDS-related deaths from 1990 to 2019, sourced from ourworldindata.org/HIV/AIDS. LR, SVM, RF and XGBoost models were introduced to predict the four-year mortality trends in South Africa and Myanmar. Model performance was evaluated using metrics including RMSE, MAE, MSE, MAPE, and Pinball. By fitting and forecasting mortality data from the past several years, it was observed that the LR model outperformed other models in predicting trends in South Africa and Myanmar. The LR model's predictions were closer to the actual outcomes. This study aids in understanding mortality trends in these regions, offers insights for epidemiological control, and helps other countries take preventive measures and formulate policies to combat this epidemic.**

Index Terms—**HIV/AIDS mortality, Time series prediction, Epidemiological insights, Machine learnin**

I. INTRODUCTION

HIV/AIDS, which stands for Acquired Immunodeficiency Syndrome, is a highly fatal infectious disease caused by the Human Immunodeficiency Virus infection [1] (the abbreviation HIV is used below). Globally, the total number of people living with HIV/AIDS in 2010 was 34 million, with an estimated increase of 3.3 million cases annually, and the highest burden in the African region[2]-[4]. It has become a global public health issue[5]-[8]. Interestingly, there are significant disparities in the HIV burden in Africa, with the sub-Saharan African region alone accounting for approximately 70% of all global HIV cases[9], [10]. Myanmar is one of the countries in Southeast Asia severely affected by the HIV epidemic[11].

The factors influencing the prevalence trends of HIV are complex and include population, economy, behavior, and environment. In recent years, scholars have used Autoregressive Integrated Moving Average (ARIMA) models, Grey System Theory models (GM (1,1)), and BP (Backpropagation) neural network models to predict the trend of HIV incidence. For example, Liang et al. [12] used the GM (1,1) model to fit the HIV incidence rate in Jiangsu Province, with a relative error of 23.89%. The GM (1,1) model is the simplest form of this model, and the basic steps in establishing it are to first accumulate irregular raw data into regular data sequences and then establish differential equations to predict future development trends of the disease[1]. Yang et al. [13] used ARIMA to establish a model for HIV incidence rates in China from 2000 to 2014, with an average absolute percentage error of 19.90%. Wu et al. [14] used Backpropagation Artificial Neural Networks (BP-ANN) as a model to predict the HIV prevalence rate, with training, calibration, and testing accuracies of 93.94%, 88.48%, and 89.60%, respectively. An et al. [15] established a two-stage Empirical Mode Decomposition (EMD) - BPNN (Backpropagation Artificial Neural Network) model to predict HIV in Dalian City. Recently, there has been considerable work on using machine learning for epidemiological dynamics prediction. For instance[17]-[19], Jr et al. [16] used a multilayer perceptron neural network to predict the time series during the nationwide outbreak of the COVID-19 pandemic. Although the full effects of COVID-19 on HIV services and development are still unknown, there are predictions suggesting that such disruptions could lead to a significant increase in HIV-related casualties.

However, there is currently limited research on predicting the trend of HIV-related death counts. This study will use four different models to predict the trend of HIV-related death counts. Figure 1 shows the mortality rate map of South Africa in 2019, while Figure 2 shows the mortality rate map of Myanmar in 2019. As seen from the figures, both South Africa and Myanmar, being high-prevalence countries, still exhibit high mortality rates. Since reporting their first cases, the HIV epidemic in both countries has been on the rise, with death counts initially increasing as well. However, in recent years, the number of deaths due to HIV in these two countries has slowly declined. Up to now, there is no cure or effective vaccine for HIV. Therefore, establishing an HIV prediction model to discover the development trend of HIV-related death counts is of great significance for HIV control efforts,

---

Manuscript received May 24, 2024; revised November 5, 2024.

This work was supported by National Natural Science Foundation of China under Grant 12001425, Natural Science Basic Research Program of Shaanxi under Grant 2023-JC-YB-066 and Shaanxi Fundamental Science Research Project for Mathematics and Physics under Grant 23JSQ049.

W. B. Yang is an Associate Professor of the School of Science, Xi'an University of Posts and Telecommunications, Xi'an 710121, China. (e-mail: yangwenbin-007@163.com).

X. Chang is a postgraduate of the School of Computer Science & Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China. (e-mail: foxchang1216@163.com).

and it can further facilitate the efficient allocation of medical aid.

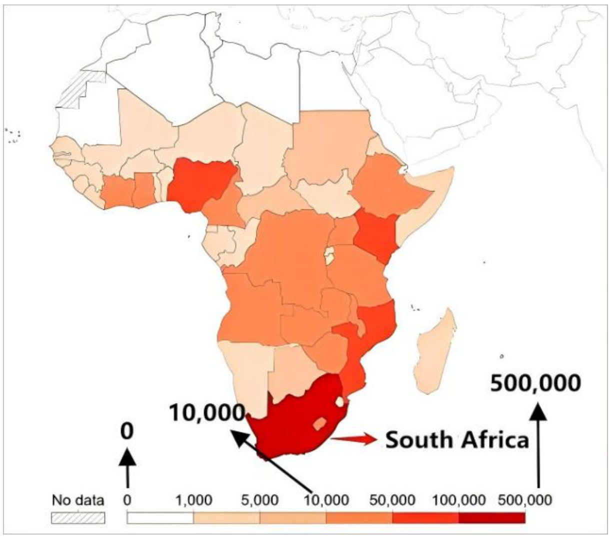

Figure 1. The mortality rate map of South Africa in 2019 (The number of deaths from HIV per 100,000 people).

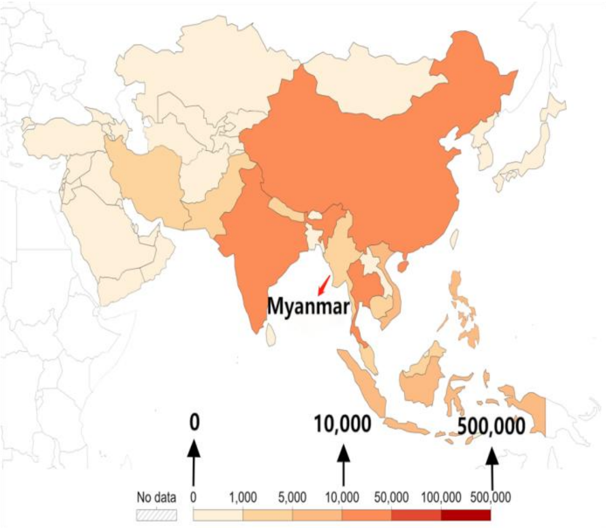

Figure 2. The mortality rate map of Myanmar in 2019 (The number of deaths from HIV per 100,000 people).

II. MATERIALS AND METHODS

A. Data Description

The data related to HIV cases in South Africa and Myanmar was collected from the Our World in Data website (HIV/AIDS - Our World in Data). This source provides comprehensive information on all HIV-related data, such as deaths, confirmed cases, and laboratory testing data. Currently, this online source is the most reliable for HIV case data. We used death case data in our study to predict the mortality trends of HIV in South Africa and Myanmar. The dataset covers data from 1990 to 2019. For training, the dataset from 1990 to 2013 was used, while the data from 2014 to 2019 was used for testing.

Figure 3 illustrates the actual trend of the number of deaths caused by the HIV virus in South Africa from 1990 to 2019. From the graph, we can observe that the number of deaths increased in a J-shaped trend from 1990 to 2006. During this period, after the end of apartheid in South Africa, there was a significant influx of foreign immigrants, which also brought many social issues such as violence, drug trafficking, and prostitution. This led to a sharp increase in HIV infections, making South Africa one of the countries with the highest HIV infection rates in the world. However, due to inadequate healthcare resources, the country couldn't meet the demands of a large number of infected individuals, resulting in an increasing number of deaths due to HIV. However, from 2006 onwards, the number of deaths started to gradually decrease. This decline is attributed to South Africa's proactive efforts in HIV detection and screening in recent years. Early detection of the HIV virus allows for early initiation of treatment, and there has been an expansion of Antiretroviral Therapy (ART) coverage, enabling more HIV-infected individuals to receive continuous treatment. ART effectively suppresses virus replication, slows disease progression, improves the quality of life, and prolongs the lifespan of patients.

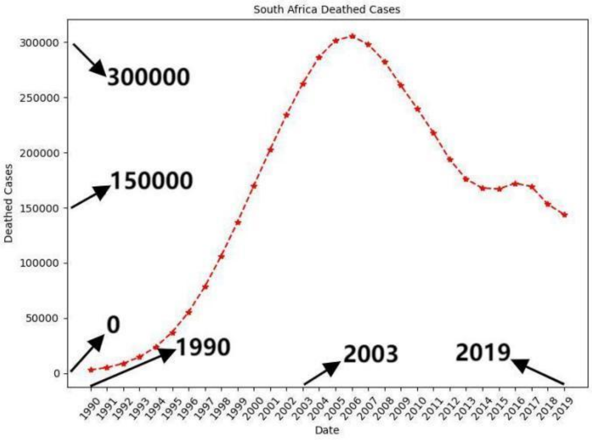

Figure 3. The trend of the number of deaths caused by the HIV virus in South Africa from 1990 to 2019.

From the Figure 4, we can observe that the number of deaths increased gradually from 1990 to 2010, and after 2010, the number of deaths began to decrease gradually.

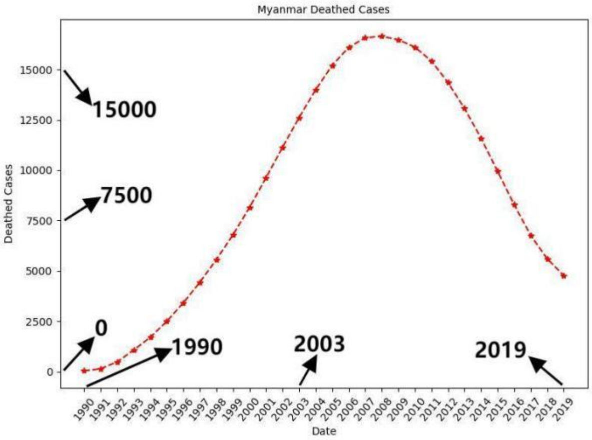

Figure 4. The trend of the number of deaths caused by the HIV virus in Myanmar from 1990 to 2019.

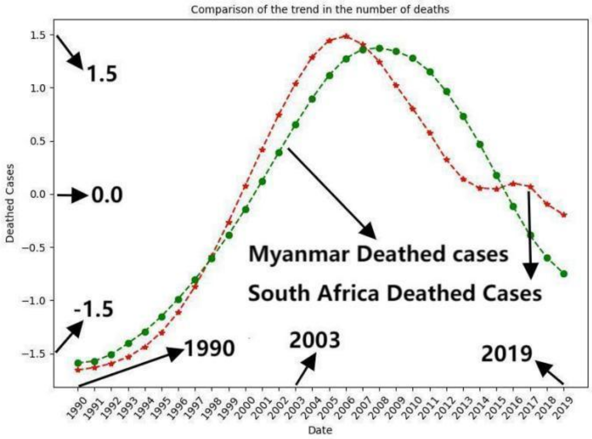

Figure 5. Comparison of the trend in the number of deaths caused by the HIV virus in South Africa and Myanmar from 1990 to 2019.

Figure 5 compares the trends in the number of deaths due to the HIV virus from 1990 to 2019 between South Africa and Myanmar. First, the data for South Africa and Myanmar is standardized to eliminate scale differences. Then, from the graphs, we can observe that the number of deaths in South Africa is significantly higher than in Myanmar. This is because South Africa is one of the countries with the highest HIV infection rates globally and has a large population of HIV-infected individuals. The high infection rate leads to more HIV-related diseases and complications, contributing to the higher number of deaths. Additionally, long-term use of antiretroviral drugs in South Africa may lead to issues of drug resistance, which can affect the effectiveness of treatment. Some regions in South Africa may also experience drug resistance, impacting the treatment outcomes for certain patients.

On the other hand, Myanmar has a smaller population, which means fewer potential transmission pathways. Moreover, Myanmar's culture, religion, and social structure may influence the spread and control of HIV. Certain cultural and social factors can affect people's behavior and attitudes, influencing the patterns of infection and treatment. International organizations, non-governmental organizations, and cooperation from other countries may have played a positive role in HIV control in Myanmar by providing technical support, drug supplies, and other resources, helping to reduce the number of deaths.

B. LR, SVM, RF and XGBoost models

Linear regression (LR) [24] is a method that transforms input variables through a linear combination to predict the target value. Linear regression offers advantages such as simplicity in form, ease of modeling, and embodies fundamental concepts in machine learning. Specifically, it constructs one or multiple models to describe the relationship between independent and dependent variables, achieved through the least squares function of the linear regression equation. As a supervised learning algorithm, it learns from a training set, deducing coefficients and subsequently evaluating the model's performance using a test set. The method often employs the least squares method, which aims to minimize the sum of squared errors, striving to find the best-fitting results. By employing the least squares method, results can be obtained quickly while minimizing the sum of squared errors between predicted and actual data points.

Support Vector Machine (SVM) [25] is a powerful supervised machine learning model primarily used for classification and regression tasks. SVM aims to find the optimal hyperplane that best separates different classes in a high-dimensional feature space. 'Support vectors' are the data points closest to the decision boundary. The effectiveness of SVM lies in its ability to handle complex data relationships and generalize well to new, unseen data. However, its performance can be influenced by the choice of kernel functions and hyperparameters, which require careful tuning. Commonly used kernel functions include linear, polynomial, radial basis function (RBF), and sigmoid kernels. SVM finds wide applications in prediction and classification tasks because it can handle complex data relationships and generalize well to new data. SVM models perform well in classifying numerical samples and are suitable for small-sample test data. The objective function is as follows:

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} \xi_i, s.t. y_i(\omega^T x_i + b) \geq 1 - \xi_i \quad (1)$$
$$\xi_i \geq 0, i = 1,2,\cdots,n,$$

where n represents the number of test samples; C is the penalty parameter $(C > 0)$; ξ represents the slack variable, ω is the coefficients of the separating hyperplane, and b is the constant parameter of the separating hyperplane. Both of these parameters need to be obtained through model training. x_i represents the i-th training sample, and y_i represents the state of the i-th sample corresponding to the path node.

The Random Forest algorithm [26] is an improvement and extension of the decision tree algorithm. It's a type of ensemble algorithm constructed based on the principles of statistical sampling. Two key parameters in this algorithm are the number of decision trees built and the number of features considered in each individual tree. The Random Forest algorithm can achieve favorable results for both classification and regression tasks. By utilizing resampling, multiple samples are drawn from the original dataset, and each sample generates its own independent decision tree. In each decision tree, a randomly selected proportion of the data is used for model building, while the remaining data is used for validation. For each sample in the test set, the model combines the results from multiple decision trees. In classification problems, it uses majority voting to make predictions, and in regression problems, it calculates the predicted values on the test set by averaging the results from multiple trees.

The XGBoost model [20]-[22] is a machine learning algorithm based on decision trees, widely used in data science. By using an internal algorithm that combines the results from multiple individual trees, XGBoost can generate accurate predictions[23]. Additionally, this model provides rankings for input features. Furthermore, XGBoost can help create stronger classifiers from other classifiers and offers other advantages such as avoiding overfitting, effectively handling missing values, and reducing runtime through parallel and distributed computing. The objective function for the XGBoost model is as follows:

$$Y^t = \sum_{i=1}^{n}(y_i, \hat{y}^{(t-1)} + f_t(x_i)) + \Omega(f_t), \quad (2)$$

where n represents the number of training data, and y_i are the feature vector and label of the i-th instance, $\hat{y}^{(t-1)}$

represents the prediction of the i-th instance at the $(t-1)$-th iteration, f_t represents the new tree that classifies the i-th instance using x_i, and represents the regularization term that penalizes the complexity of the new tree.

C. Models performance measures

To evaluate the performance of each of the models mentioned above, commonly used accuracy measurement methods were applied. The performance functions are as follows [27]:

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{n}\frac{|\tilde{\omega}_i - \omega_i|}{\omega_i}; \tag{3}$$

Mean absolute error (MAE):

$$MAE = \frac{1}{N}\sum_{i=1}^{n}|\tilde{\omega}_i - \omega_i|; \tag{4}$$

Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(\omega_i - \tilde{\omega}_i)^2}; \tag{5}$$

Mean Square Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(\omega_i - \tilde{\omega}_i)^2; \tag{6}$$

Pinball:

$$L_\tau(y,z) = \begin{cases}(\omega_i - \tilde{\omega}_i)\tau, & \omega_i \geq \tilde{\omega}_i, \\ (\omega_i - \tilde{\omega}_i)(1-\tau), & \omega_i < \tilde{\omega}_i.\end{cases} \tag{7}$$

In these above equations (3)-(7), N, ω_i and $\tilde{\omega}_i$ represent the number of observations, actual values, and predicted values, respectively, while $\omega_i - \tilde{\omega}_i$ represents the error between predicted values and actual values. MAE is the arithmetic mean of the absolute errors between predicted values and true values, providing the average of absolute prediction errors. MSE is a loss function used to measure the error between predicted values and true values. RMSE is the square root of the mean squared error and is often used to assess the differences between predicted values and true values. MAPE is expressed as a percentage and represents the average error between actual values and predicted values, calculating the average absolute percentage difference between actual and predicted values. Pinball is also a loss function used to evaluate the accuracy of quantile forecasts, where τ represents the target quantile.

When MAPE, RMSE, MAE, MSE, and Pinball values approach zero, the prediction results are considered more accurate.

III. RESULTS

In this article, trends in the number of deaths caused by HIV in South Africa and Myanmar were predicted using SVM, LR, Random Forest, and XGBoost. The data was divided into two parts: one for training and the other for testing the models. The data was split into a training set and a testing set in an 80:20 ratio. After training, the models were used to predict the number of deaths using the testing set. Figure 6 represents the predictions for South Africa's testing data, while Figure 7 represents the predictions for Myanmar's testing data. The blue curve represents the data from the testing set, and the red curve represents the predicted data using different models. The learning results are shown in the figures. From Figures 6 and 7, we can roughly observe that the LR model yields better predictions compared to the other three models.

Figure 8 presents a comparison of the death trends predicted by different models for South Africa, while Figure 9 shows the comparison for Myanmar. These figures display the predictions of all four models alongside the actual test data for HIV-related deaths, providing a more intuitive view of each model's prediction performance. The blue line represents the actual test data, the orange line represents the SVM model's predictions, the green line represents the Random Forest model's predictions, the red line represents the Linear Regression model's predictions, and the purple line represents the XGBoost model's predictions. The comparison results are more clearly displayed here, with the Linear Regression model having predictions closest to the actual values, followed by the SVM model, Random Forest model, and XGBoost model.

The prediction results for the number of HIV-related deaths in South Africa show that the LR model has an MAE of 0.0374, an MSE of 0.0014, an RMSE of 0.0383, an MAPE of 18.3070%, and a Pinball of 0.0187, which are better than the evaluation metrics of the other three models. The specific prediction results evaluation metrics for SVM, LR, Random Forest, and XGBoost are shown in Table I. For the prediction results of deaths in Myanmar, the LR model has an MAE of 0.0203, an MSE of 0.0005, an RMSE of 0.0219, a MAPE of 18.7549%, and a Pinball of 0.0101, which is superior to the other three models. The evaluation metrics for the prediction results of other models are shown in Table II.

The prediction results for the number of HIV-related deaths in South Africa show that the LR model has an MAE of 0.0374, an MSE of 0.0014, an RMSE of 0.0383, an MAPE of 18.3070%, and a Pinball of 0.0187, which are better than the evaluation metrics of the other three models. The specific prediction results evaluation metrics for SVM, LR, Random Forest, and XGBoost are shown in Table I. For the prediction results of deaths in Myanmar, the LR model has an MAE of 0.0203, an MSE of 0.0005, an RMSE of 0.0219, a MAPE of 18.7549%, and a Pinball of 0.0101, which is superior to the other three models. The evaluation metrics for the prediction results of other models are shown in Table II.

The LR model is often used as a baseline model for comparison with other more complex models. In some cases, the performance of a linear model may unexpectedly surpass that of other models. The overall trend of HIV in South Africa is not linear, but the prediction results indicate that the linear model performs better than other models. This could be because certain local regions of the data exhibit a linear relationship, and the linear model may better fit the data in these areas. Additionally, the overall data size is relatively small, making it easier for the linear model to achieve a more robust performance. This result validates why the LR model is often used as a baseline for comparison with more complex models. In some cases, the performance of a linear model may unexpectedly outperform other models.

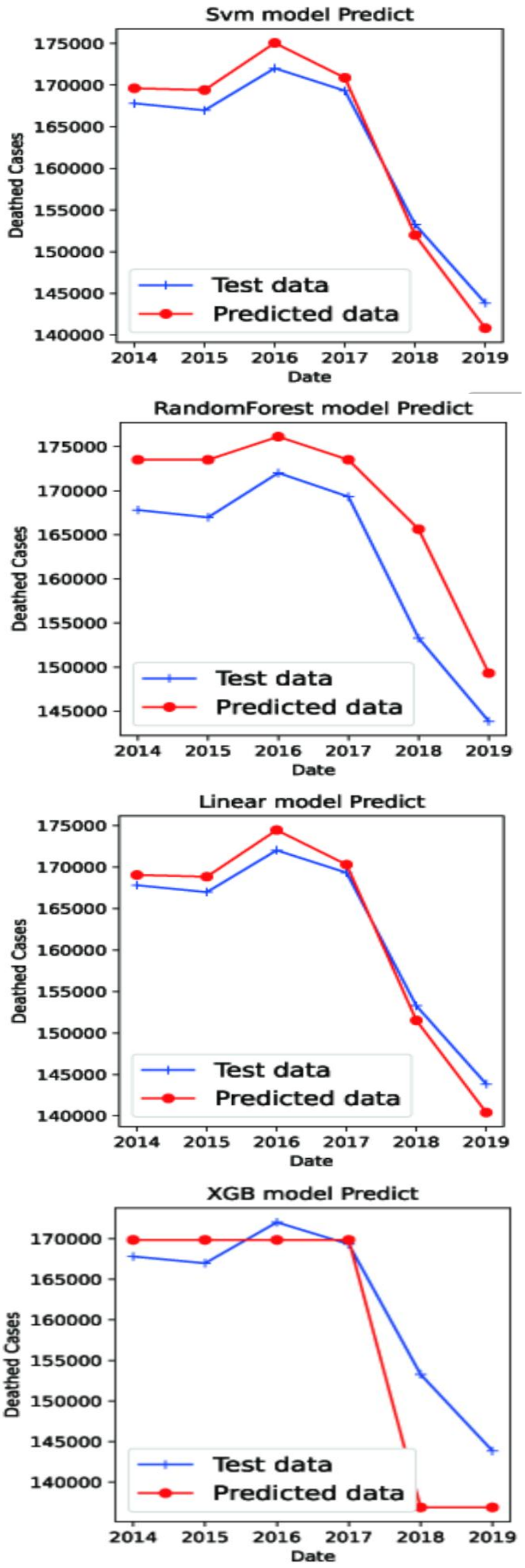

Figure 6. Prediction chart for South African test data (red line). The blue line represents the actual values of test set.

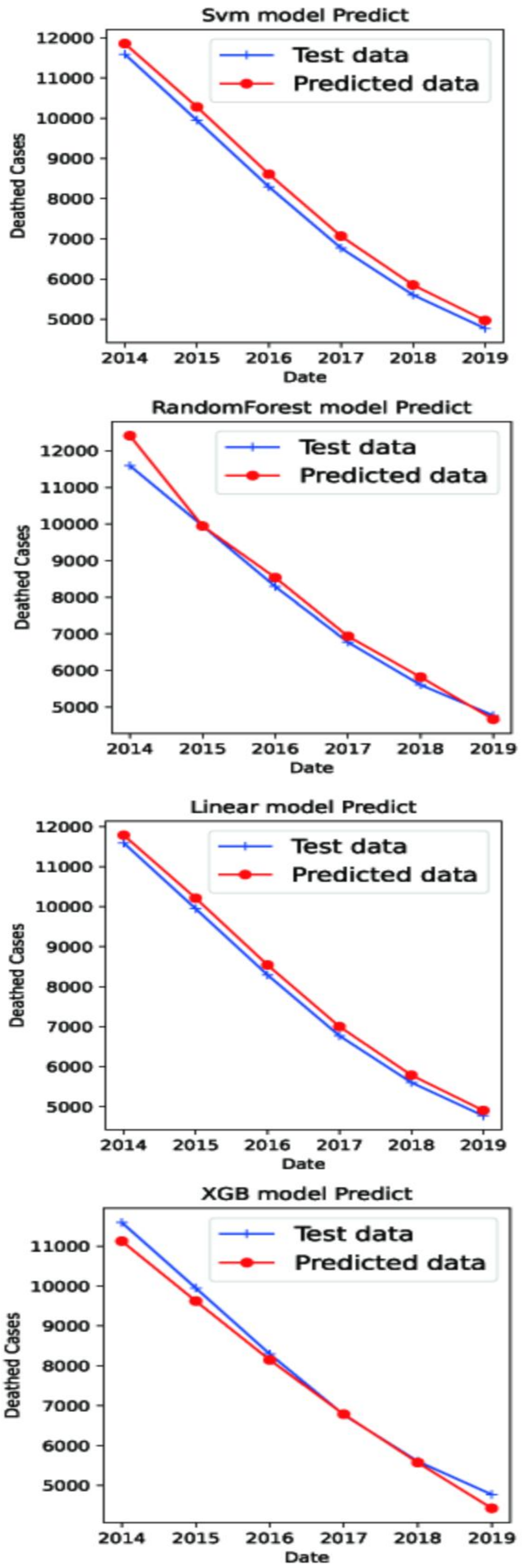

Figure 7. Prediction chart for Myanmar test data (red line). The blue line represents the actual values of test set.

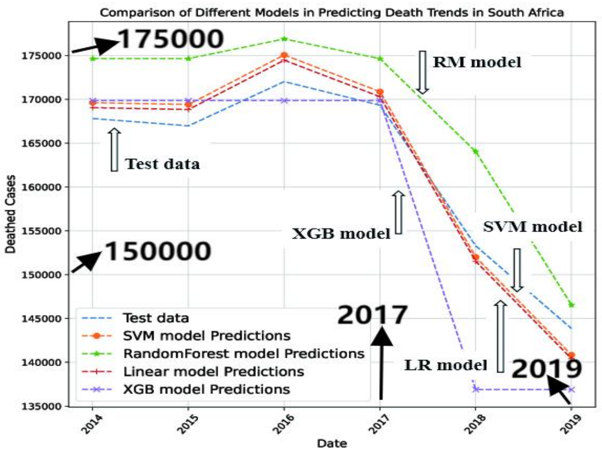

Figure 8. Comparison of different learning models in predicting death trends in South Africa.

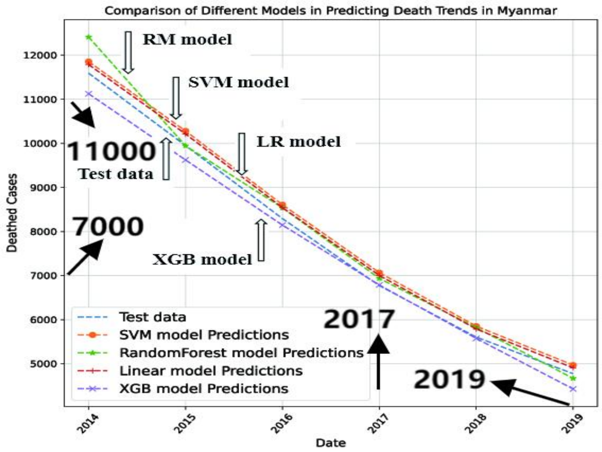

Figure 9. Comparison of different learning models in predicting death trends in Myanmar.

IV. DISCUSSION

Figure 10 presents the trend prediction of HIV mortality in South Africa from 2020 to 2023. The blue line represents historical actual data, while the red line indicates future forecast data. It shows that over the next four years, the annual number of deaths is expected to decrease year by year, exhibiting a general downward trend, though the decline is relatively slow. The figure also provides the estimated range of HIV deaths, with the yellow line representing the estimated maximum number of deaths and the green line representing the estimated minimum. The data is sourced from the AIDSinfo website, with the forecast data extending up to the end of 2022. Figure 11 depicts the trend prediction of HIV mortality in Myanmar from 2020 to 2023, with the blue line illustrating historical actual data, the red line future forecast data, the yellow line the estimated maximum number of deaths, and the green line the estimated minimum. Over the next four years, the annual number of deaths in Myanmar is also predicted to decrease, but at a faster rate compared to the trend in South Africa.

Regarding the forecast data for South Africa, this study's predictions do not fall within the official forecast range. This discrepancy is attributed to the official estimates of mortality being generally lower since 2012, likely due to the development of antiretroviral drugs and the widespread availability of treatment, leading to overly optimistic official mortality estimates. However, due to a large base of infected individuals and the time required for treatment to become widely accessible, this study's predictions based on actual mortality rates show a slight deviation from official estimates, although the overall trend remains similarly positive. For Myanmar, the study's forecast data falls within the official prediction range. Despite actual mortality rates being significantly higher than the forecast range before 2014, the conservative nature of the forecasts, the growing forecast range, and increased investments in HIV treatment in Myanmar have led to a gradual and slow decrease in death rates, falling back within the estimated range. Under the current conditions of disease prevention measures, social environment, and investments in medical services, both South Africa and Myanmar are experiencing a positive development trend with decreasing HIV mortality rates.

TABLE I
COMPARING THE PERFORMANCE OF DIFFERENT LEARNING MODELS FOR PREDICTING THE TREND OF DEATH CAUSED BY HIV IN SOUTH AFRICA

Model	Model fitting				
	MSE	MAE	RMSE	MAPE	Pinball
SVM	0.0024	0.0492	0.0490	25.8271%	0.0246
LR	0.0014	0.0374	0.0383	18.3070%	0.0187
RM	0.0040	0.0500	0.0637	19.2065%	0.0250
XGBoost	0.0067	0.0564	0.0823	32.4223%	0.0282

TABLE II
COMPARING THE PERFORMANCE OF DIFFERENT LEARNING MODELS FOR PREDICTING THE TREND OF DEATH CAUSED BY HIV IN MYANMAR

Model	Model fitting				
	MSE	MAE	RMSE	MAPE	Pinball
SVM	0.0006	0.0227	0.0238	21.5210%	0.0113
LR	0.0005	0.0203	0.0219	18.7549%	0.0101
RM	0.0051	0.0678	0.0712	152.7885 %	0.0339
XGBoost	0.0062	0.0561	0.0787	58.9220%	0.0281

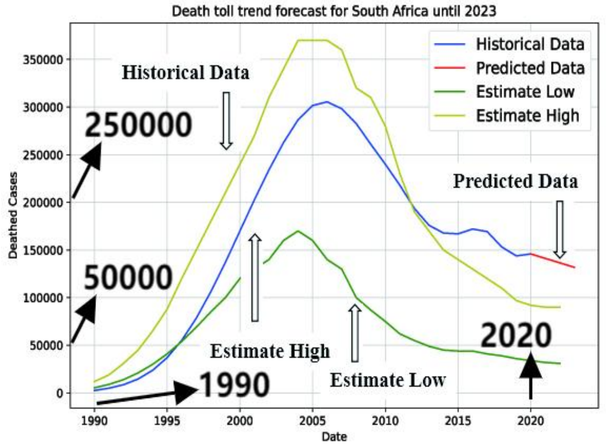

Figure. 10. Death toll trend forecast for South Africa until 2023 (red line).

Note that HIV is one of the deadliest infectious diseases in the world, causing nearly 1 million deaths annually. In some countries, particularly in the South African region, HIV is the leading cause of death. With the increasing availability of treatment options for HIV, the use of machine learning for time series forecasting has proven effective in modeling and predicting the future trajectory of virus transmission.

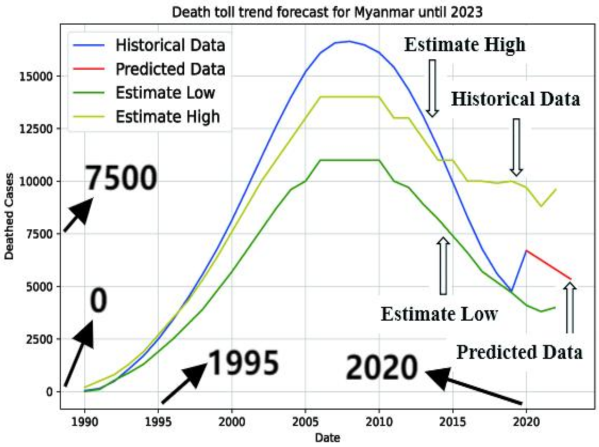

Figure. 11. Death toll trend forecast for Myanmar until 2023 (red line).

In this study, we introduced machine learning models, including SVM, LR, Random Forest, and XGBoost, to predict the trends in HIV-induced deaths in South Africa and Myanmar. To train and test the models used for our research, we utilized data up to the year 2019. The models employed in this work are data-driven approaches, and we assessed our model predictions using MAE, MSE, RMSE, MAPE, and Pinball. Using the models we developed, we forecasted the number of deaths for the next four years. Our findings contribute to the more rational allocation of HIV medical resources, prioritizing assistance in countries with persistently high death rates. For instance, this can enhance antiretroviral therapy programs in regions with high mortality rates, ensuring their accessibility and delivery. This can aid these areas in implementing preventive measures against the epidemic and formulating policies to reduce mortality rates.

V. LIMITATIONS

The research area of this article, South Africa and Myanmar, has a relatively small amount of data, covering only the mortality data from 1990 to 2019. Official websites do not provide mortality data from 2020 to the present. However, we can indirectly corroborate our predictive results through the estimated mortality data range provided by official sources up to the end of 2022. Additionally, the data from the test set can be used to validate the training results of the model, showing that the LR model performs better. Although the overall trend of HIV in South Africa is not linear, some local areas of the data exhibit linear relationships, and linear models may better fit the data within these areas. With the overall smaller data volume, linear models are more likely to achieve robust results. If a larger dataset were available, a non-linear trend might emerge, potentially leading to better predictive performance from models like Random Forest, Support Vector Regression, and XGBoost.

REFERENCES

[1] R. A. Weiss, “How does HIV cause AIDS?,” Science, vol. 260, no. 5112, pp. 1273-1279, 1993.

[2] “Global HIV/AIDS Response: Epidemic update and health sector progress towards Universal Access, Progress Report 2011,” WHO Africa Publications, 2012.

[3] Global HIV & AIDS statistics - Fact sheet, UNAIDS, 2022.

[4] K. Zuma, L. Simbayi, N. Zungu, “The HIV epidemic in South Africa: Key findings from 2017 national population-based survey,” International Journal of Environmental Research and Public Health, vol. 19, no. 13, pp. 8125, 2022.

[5] T. D. Frank, A. Carter, D. Jahagirdar, “Global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2017, and forecasts to 2030, for 195 countries and territories: A systematic analysis for the Global Burden of Diseases, Injuries, and Risk Factors Study 2017,” Lancet HIV, vol. 6, no. 12, pp. e831-e859, 2019.

[6] “Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results,” Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.

[7] P. Haldar, S. Reza-Paul, R. A. Daniel, “A rapid review of pre-exposure prophylaxis for HIV in the Asia-Pacific region: recommendations for scale up and future directions,” Sexual Health, vol. 18, no. 1, pp. 31-40, 2021.

[8] Y. Yang, S. Mao, Y. Xue, “Prediction on the incidence rate of AIDS in China with GM(1, 1) and trend extrapolation model,” Chinese Journal of Health Statistics, vol. 31, pp. 952-954, 2014.

[9] UNAIDS, AIDSinfo, 2022.

[10] T. Rehle, L. Johnson, T. Hallett, “A Comparison of South African National HIV Incidence Estimates: A Critical Appraisal of Different Methods,” PLoS ONE, vol. 10, no. 7, pp. e0133255, 2015.

[11] B. Williams, D. Baker, M. Bühler, “Increase coverage of HIV and AIDS services in Myanmar,” Conflict and Health, vol. 2, pp. 3, 2008.

[12] Y. Liang, Y. Wang, L. Zhai, “Analysis of forecasting the incidence of AIDS, gonorrhea, and syphilis in Jiangsu Province with GM(1, 1) model,” Journal of Nantong University (Medical Sciences), vol. 1, pp. 21-24, 2013.

[13] Q. Yang, G. Guo, H. Chen, “Research on AIDS incidence prediction based on ARIMA,” China Medical Equipment, vol. 13, pp. 1-4, 2016.

[14] S. Meakin, S. Abbott, N. Bosse, “BP-neural network as a model of predicting STD/AIDS prevalence,” Chinese Journal of AIDS & STD, vol. 6, pp. 525-528, 2007.

[15] Q. An, J. Wu, J. Meng, “Using the hybrid EMD-BPNN model to predict the incidence of HIV in Dalian, Liaoning Province, China, 2004-2018,” BMC Infectious Diseases, vol. 22, no. 1, pp. 102, 2022.

[16] S. A. Jr, G. B. De, R. Medina, “Time Series Forecasting of HIV/AIDS in the Philippines Using Deep Learning: Does COVID-19 Epidemic Matter?,” International Journal of Emerging Technology and Advanced Engineering, vol. 12, no. 9, pp. 144-157, 2022.

[17] C. Mweemba, P. Hangoma, I. Fwemba, “Estimating district HIV prevalence in Zambia using small-area estimation methods (SAE),” Population Health Metrics, vol. 20, no. 1, pp. 8, 2022.

[18] Y. Wang, Z. Yan, D. Wang, “Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models,” BMC Infectious Diseases, vol. 22, no. 1, pp. 495, 2022.

[19] A. K. Gupta, V. Singh, P. Mathur, “Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet, and linear regression models in Indian scenario,” Journal of Interdisciplinary Mathematics, vol. 24, no. 1, pp. 89-108, 2021.

[20] K. Menguc, N. Aydin, A. Yilmaz, “A data-driven approach to forecasting traffic speed classes using the Extreme Gradient Boosting algorithm and graph theory,” Physica A: Statistical Mechanics and its Applications, vol. 620, pp. 128738, 2023.

[21] J. Luo, Z. Zhang, Y. Fu, “Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms,” Results in Physics, vol. 27, pp. 104462, 2021.

[22] Z. Fang, S. Yang, C. Lv, “Application of a data-driven XGBoost model for the prediction of COVID-19 in the USA: a time-series study,” BMJ Open, vol. 12, no. 7, pp. e056685, 2022.

[23] M. Mehta, J. Julaiti, P. Griffin, “Early Stage Machine Learning Based Prediction of US County Vulnerability to the COVID-19 Pandemic (Preprint),” JMIR Public Health and Surveillance, vol. 6, no. 3, pp. e19446, 2020.

[24] S. Rath, A. Tripathy, A. R. Tripathy, “Prediction of new active cases of coronavirus disease (COVID-19) pandemic using a multiple linear regression model,” Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 5, pp. 1467-1474, 2020.

[25] Y. Tian, C. Chen, X. Chen, “Research on real-time analysis technology of urban land use based on support vector machine,” Pattern Recognition Letters, vol. 133, pp. 320-326, 2020.

[26] Y. Wang, D. Wang, L. Zhang, “Automatic identification of rapid eye movement sleep based on random forest using heart rate variability,” Physica A: Statistical Mechanics and its Applications, vol. 527, pp. 121421, 2019.

[27] C. Barria-Sandoval, G. Ferreira, K. Benz-Parra, “Prediction of confirmed and death cases of Covid-19 in Chile through time series techniques: A comparative study,” PLoS One, vol. 16, no. 4, pp. e0245414, 2021.