



## Performance Comparison of Deep Learning Models in Predicting HIV Incidences in Tanzania

Zaituni Kaijage\*, Honest Kimaro and Hellen Maziku

*College of Information and Communication Technologies, University of Dar es Salaam,*

*P. O. Box 33335, Dar es Salaam, Tanzania;*

*\*zaykaijage@gmail.com (ZAK)*

*\*Corresponding author*

Received 10 Feb 2024, Revised 2 May 2024, Accepted 25 Jun 2024, Published 30 Jun 2024

<https://dx.doi.org/10.4314/tjs.v50i2.15>

### Abstract

Human immunodeficiency virus infection/acquired immune deficiency syndrome (HIV/AIDS) is a global pandemic that has claimed more than 40 million lives since it was discovered in the late 1970s. In sub-Saharan Africa, including Tanzania. Different measures to combat the diseases have failed to be attained, like the UNAIDS 90-90-90 target, which aimed to reduce HIV by 2020, and it was moved to 2030. The availability of proper tools to control and monitor diseases and ensure proper early intervention is very important. Prediction of disease trends using Machine Learning (ML) models can improve speed towards attaining the UNAIDS targets by providing accurate insights into the disease trends. The performance of ML models depends on many factors, including datasets that influence the generalization of models. This study aims to suggest the best deep-learning model to predict HIV incidences in Tanzania. Four deep learning models, recurrent neural network (RNN), Gated Recurrent unit (GRU), Long Short-Term Memory (LSTM), and 2D convolution layer (CONV2D), have been studied. HIV data is collected from District Health Information System 2 (DHIS2), the national Health Management Information System (HMIS). The HIV data collected is for 26 regions in Mainland Tanzania, recorded from January 2015 to October 2022. The accuracy of the models was evaluated using three metrics: Mean absolute error (MAE), Mean absolute Percentage Error (MAPE), and mean square error (MSE). The results show that Conv2D achieved the lowest average training time for short-term predictions, while RNN records the highest accuracy with the lowest MAE for all considered cases. The GRU was the fastest for the long-term predictions, and the LSTM reported the best accuracy.

**Keywords:** AI Models; Algorithms; Deep learning models; HIV incidence; Machine Learning; Prediction

### Introduction

HIV is a global pandemic that has claimed more than 40.1 million people globally, and more than 38.4 million people are living with HIV (UNAIDS 2022). HIV causes acquired immunodeficiency syndrome (AIDS), which results in T cells not functioning correctly (Li et al. 2020). According to the report published by the World Health Organization (WHO) in November 2022, there were 1.5 million

people who became newly infected with HIV in 2021, and 650,000 people died from AIDS-related illnesses in 2021 (UNAIDS 2022, WHO 2022). In Tanzania, prevalence in the adult population was estimated to be 6.3% for females and 3.4% among men by 2023 (UNAIDS 2020). Different initiatives have been established to end HIV/AIDS, including the 90-90-90 target, which aimed at making sure that 90% of all people living with HIV will know their HIV status, 90% of

all people diagnosed with HIV infection will receive sustained treatment, and 90% of all people receiving treatment will have viral load suppression. However, the target failed and was moved to 2030 (WHO 2022). Despite the efforts set by international organizations, HIV/AIDS is still a public threat and a source of many deaths (Wang et al. 2019a). Different methods can be used to control and monitor the diseases for early interventions. One of the methods is predicting the future trend of the diseases and imposing proper early interventions for disease control (Aribe et al. 2022). The disease prediction results can be used by stakeholders, government, policymakers, and health professionals to design and make interventions and future planning. Studies have shown that traditional statistical and machine learning (ML) algorithm techniques have been used to predict future trajectory and distribution of diseases (Dansana et al. 2020; Stavelin 2021). However, statistical models cannot capture hidden nonlinear features and are inaccurate (Elsaraiti and Merabet 2021). Machine learning allows systems to learn from experience and automatically determine hidden data associations that increase prediction accuracy (Sarker 2021). In HIV interventions, ML has been used in various cases, such as identifying HIV predictors (Mutai et al. 2021), predicting HIV status (Ahlström et al. 2019), and identifying patients at risk of HIV acquisition (Marcus et al. 2019). In developing countries, researchers have developed classical ML models for solving HIV/AIDS issues such as Uganda (Balzer et al. 2020), sub-Saharan Africa (Mutai et al. 2021), Rwanda (Nyoni and Nyoni 2021b), Kenya (Nyoni and Nyoni 2021a) and Tanzania (Chikusi and Kaijage 2022). While the researchers have applied a wide range of classical ML algorithms, such as random forest (RF), XGBoost, and artificial neural networks (ANN), little work has been done using deep learning (DL) models. Studies show that, in predicting HIV incidences, deep learning models are superior to statistical and classical ML models (Alehegn 2022; Li et al. 2020; Wang et al. 2019b). In China, (Wang et

al. 2019a) developed a DL model to describe the tendency of HIV and predict its incidence. However, their model could not be generalized for predictions in other countries (Aribe et al. 2022). The most common DL models for time-series prediction are standard RNN, Conv2D, GRU, and LSTM (Baccouche et al. 2020). Different studies conducted have shown DL models perform differently depending on the dataset and other model parameters (Wang et al. 2023). For instance, the DL models, such as LSTM, were used to predict different diseases, including coronavirus, and it outperformed another model (Arun Kumar et al. 2022; Obaid et al. 2020) heart diseases (Baccouche et al. 2020), Lung diseases Jasmine et al. (2023) and tuberculosis (Yang et al. 2022). Despite the excellent performance of DL models in predicting diseases, few studies have been conducted in predicting HIV incidences, specifically in developing countries like Tanzania. There are various DL studies conducted in China (Wang et al. 2019a), in the Philippines (Aribe et al. 2022), and in the United States (Jiang et al. 2022), but these studies were based on developed countries with different geographical settings, which may not be efficient in developing countries like Tanzania, and can result in low prediction accuracy since effective HIV interventions are not generic (Kagaayi et al. (2019). In this study, Four DL models, including LSTM, GRU, Standard RNN, and Conv2D models, were used to predict HIV incidence in Tanzania, and the results were analyzed to identify which model is the best and most accurate for local HIV incidence prediction in Tanzania for proper interventions.

## **Literature review**

### **HIV/AIDS Prediction and Deep Learning**

More than 90% of newly diagnosed HIV cases worldwide are attributed to Tanzania, in addition to 34 other nations. The country has implemented a response strategy to mitigate the prevalence of the HIV/AIDS epidemic and enhance the well-being of those affected ever since its entrance in the early 80s (Lekey-Kawo, 2023). Through its UNAIDS,

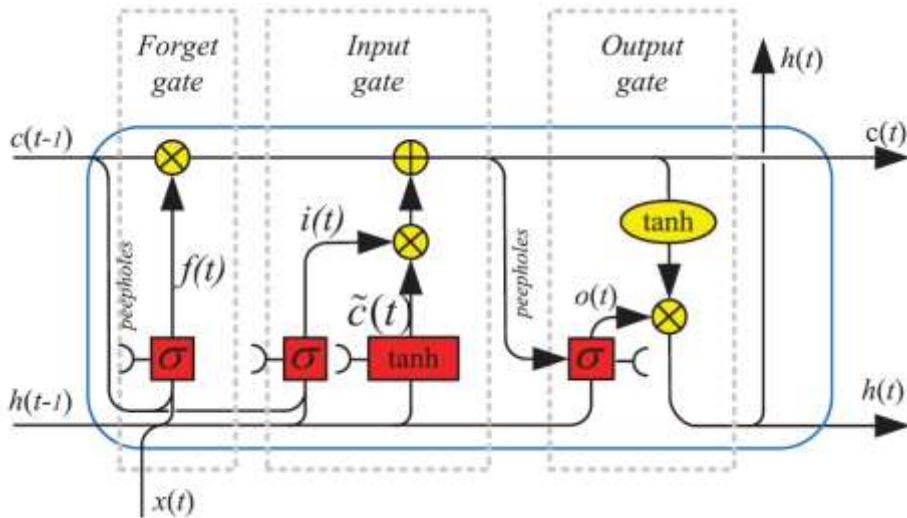
the United Nations has established different strategies to control and reduce disease. One of the notable initiatives was establishing a 90-90-90 target, which aimed at ending HIV spread by 2020 (CDC, 2019), but the target failed and was shifted to 2030. According to the World Health Organisation (WHO), disease prediction is one method that can effectively control and monitor HIV/AIDS disease (Wang et al. 2019a). Accurate disease predictions can help the government and stakeholders formulate targeted interventions for prevention and control (Ahlström et al. 2019). HIV prediction can be achieved through AI algorithms, and DL learning algorithms are one of the AI prediction algorithms (Alehegn, 2022). Deep learning models are sophisticated artificial neural networks (ANN) with many trainable parameters. These models can perform complex tasks exceptionally, including computer vision, natural language processing, image, and sound data classification, and predictions.

### **The standard RNN, GRU, LSTM, and Conv2D models**

Recurrent Neural Networks (RNN) work on saving the output of a particular layer and feeding this back to the input to predict the layer's output. RNNs are networks of standard recurrent cells such as sigma and tanh cells. Recurrent neural network (RNN) models can capture hidden nonlinear features and provide good results for disease

forecasting. The most common deep learning models used for time-series prediction are Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) (Baccouche et al. 2020; Obaid et al. 2020; Shi and Zhang 2022). Recently, Shi and Zhang (2022) proposed the deep learning model by combining a convolutional neural network (CNN) and GRU to predict hypertension. The popularity of GRU comes from its simple structure and relatively computational time efficiency. However, the GRU suffers from the vanishing gradient problem, especially when the timestamp is significant. Therefore, the GRU is only efficient for forecasting small timestamps (Yamak et al. 2019).

Hochreiter and Schmidhuber (1997) introduced LSTM to deal with the vanishing gradient problem in RNN and GRU. They improved the remembering capacity of the standard recurrent cell by introducing a "gate" into the cell. Since this pioneering work, LSTMs have been modified to several variations, such as LSTM without a forget gate, LSTM with a forget gate and LSTM with a peephole connection. Usually, the term LSTM cell denotes LSTM with a forget gate, as shown in Figure 1. The LSTM has been used for the forecasting of many diseases such as coronavirus (Obaid et al. 2020), heart diseases (Baccouche et al. 2020), hand-foot-mouth diseases (Gu et al. 2019), tuberculosis (Yang et al. 2022) and HIV (Baytas et al. 2017).



**Figure 1:** Architecture of LSTM with a forget gate (Yu et al. 2019).

A memory cell in an LSTM unit can store data for long periods, and three gates manage the flow of information into and out of the cell. For instance, the ‘Forget Gate’ determines what information from the previous state cell should be memorized and what information no longer useful should be removed. The ‘Input Gate’ determines which information should enter the cell state, and the ‘Output Gate’ determines and controls the outputs (Van Houdt et al. 2020). As it solves the issues of training a recurrent network, the LSTM network is considered one of the most successful RNNs. Although variants and analogous neural cells have been introduced, they can only be used on one or some specific data sets. Additionally, no cell variant can outperform the LSTM cell overall. Thus, the LSTM cell is still the focus of deep learning for processing sequential data, and it remains the most popular recurrent cell in the literature (Yu et al. 2019).

The LSTM is designed to work with data with regular patterns. In healthcare systems with EHR data that are irregular and temporal, the conventional LSTM models may not work efficiently (Baytas et al. 2017). Traditionally, LSTM adopts an offline batch mode for model training. To be updated with new data, the network must be re-trained with merged data using both old and new data, which is very time-consuming and causes

catastrophic forgetting (Wang et al. 2021). LSTM has been combined with different methods in healthcare systems, such as the aware technique and model decomposition, to develop a hybrid model (Xia et al. 2019). In comparing the performances, the hybrid LSTM methods outperform the other single DL methods (Nosratabadi et al. 2020). Also, in other applications, the LSTM has been combined with Model Decomposition techniques to reduce time complexity and improve performance (Chen et al. 2019; Jaseena and Koor 2021).

### Related Works

Deep learning can leverage labelled datasets to observe patterns in the data and cluster inputs appropriately with no or very little human intervention (Sarker 2021). As deep learning models designed to capture long dependencies in sequential data, the LSTM has been widely used for predicting incidents and prevalence of diseases such as tuberculosis, diabetes mellitus, COVID-19 and HIV (Ceia-Hasse et al. 2023; Yang et al. 2022). Wang et al. (2019b) developed a model to describe the tendency of HIV and to forecast its incidence using LSTM. The authors used HIV incidence data from Guangxi province, China, stored in EHR from 2005 to 2016. The LSTM model was more effective than other time series models,

such as ARIMA and Neural Networks (NN). Aribé et al. (2022) proposed an ANN model for country-wise predictions of HIV epidemics and the impact of Antiretroviral Therapies (ART) until 2030. They used monthly datasets extracted from the HIV/AIDS and ART registry of the

Philippines from March 2009 – February 2022. Their study used the most recent datasets and provided country-wise predictions. However, they did not consider the LSTM model, which is regarded as the most powerful prediction model.

**Table 1:** Machine Learning Models Studies for Predicting HIV Incidents

SN	Authors	Models Accuracy	Datasets	Remarks
1	Wang et al. (2019b)	LSTM	EHR data from 2005 to 2016 (China)	A single province was involved. More recent datasets may be used.
2	(Aribé et al. 2022)	ANN	EHR data from March 2009 – February 2022 (Philippines)	They used ANN. More efficient prediction models may be used (LSTM)
3	(Zhang et al. 2022)	CNN, RNN, and LSTM	Cohort Study from 2005 to 2020 (United States)	A single state was involved. Based on the Cohort study
4	(Alehegn 2022)	ML, GRU, and LSTM	Worldwide data collected by California Health and Human Services	Not the case for HIV interventions

Another study by (Zhang et al. 2022) developed a model for examining the dynamic patterns of viral suppression and predicting the various viral suppression indicators. They translated the models to a beta version of service-ready tools for clinical decision support. They used a statewide cohort of people living with HIV in South Carolina, United States, from 2005 to 2020. In their study, they developed CNN, RNN, and LSTM-based models. However, according to their results, the authors could not conclude which model is the best. The study was based on a cohort that may be expensive and sometimes provide biased data. Although they considered county-level factors, their model was developed for only one state. Alehegn (2022) compared ML models, GRU, and LSTM for predicting HIV/AIDS and reported that LSTM models are better than ML models. However, general worldwide datasets were used, which may not

be applicable in every context because key determinants of HIV interventions vary from one society to another (Kagaayi et al. 2019).

Table 1 presents the most recent DL models for predicting HIV incidents. The authors have shown the DL models for predicting HIV in different countries. However, HIV interventions depend on the behaviors and socioeconomic status of society. In Africa, a combination of HIV interventions, including antiretroviral therapy (ART), voluntary medical male circumcision, HIV testing services, the use of condoms, and behavior change interventions, have been shown to reduce HIV incidence in some lower-risk populations (Kagaayi et al. 2019). Therefore, since ML models are data-driven, studies performed in developed countries or far different geographical areas, like those presented in Table 1, may not be efficient for a country like Tanzania (Kagaayi et al. 2019). Several researchers have developed ML

models for dealing with HIV/AIDS by looking into different perspectives in African countries. (Mutai et al. 2021) proposed an ML model for identifying HIV predictors and predicting people at high risk of infections, specifically in sub-Saharan Africa. The study used data from fourteen countries, including Tanzania, Kenya, Uganda, Rwanda, Zambia, and Malawi. In Combating HIV, ARTs have been used to reduce the side effects of HIV; predicting the number of people enrolled in ART can provide predictions on HIV prevalence. Studies have been done to predict the coverage of ARTs and HIV prevalence in different countries, such as Rwanda (Nyoni and Nyoni 2021b) and Kenya (Nyoni and Nyoni 2021a). In (Nyoni and Nyoni 2021a), the authors used a Multilayer Perceptron Neural Network to analyze ART coverage in Kenya. In Tanzania, an ML model was developed by Chikusi and Kaijage (2022) to predict and visualize HIV index testing. The developed model used the data collected from Tanzania's Kilimanjaro, Arusha, and Manyara regions. Another ML model for identifying persons at high risk of HIV acquisition was proposed by (Balzer et al. 2020). The authors used population-level HIV testing data from rural Kenya and Uganda to construct HIV risk scores. The presented HIV intervention strategies using artificial intelligence in Sub-Saharan Africa are mainly based on classical ML models with poor performances. Also, none or only a few studies of the reported studies focus on the prediction of new HIV cases despite many victims. More studies on the LSTM DL technique should be conducted for better HIV forecasting results to help the government create proper interventions for HIV/AIDS control.

### **Performance Evaluation Techniques.**

The accuracy of trend results measures the degree of difference between forecasting results and actual data. Forecasting inaccuracy (error) can be measured by deviation and bias. However, ideally, prediction techniques produce zero deviation and zero bias. This study used MSE, MAE, and MAPE to compare the algorithms'

performances. Many research studies have applied these indicators to determine the best forecasting model. Small measurement values indicate a more accurate estimate as they result in a minimum approximate error. The RMSE is a good measure of accuracy, but only to compare forecasting errors of different models or model configurations for a particular variable, not between variables, as it is scale dependent. The equation for RMSE is presented in (1), and MAPE is presented in (2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right| \quad (2)$$

where  $y_i$  is the prediction value,  $x_i$  the true values and  $n$  is the number of observations.

## **Materials and Methods**

### **Data Sources**

The data was collected from the District Health Information Software 2 (DHIS2) platform. The platform has tools for collecting, validating, analyzing, and presenting aggregate statistical data tailored to integrated health information management activities. In Tanzania, the system is currently adopted and used by the Ministry of Health (MoH) as a nationwide HMIS data warehouse (DHIS2 2023). HIV data from January 2015 to October 2022 for all Tanzania mainland regions was collected and used to forecast new HIV cases in the country. The dataset contains the new monthly HIV cases for each region. This study involves experimental research whereby the collected data from DHIS2 were used to predict the HIV prevalence in Tanzania using deep learning models. It should be mentioned that all available data in the DHIS2 were used therefore data sampling was performed.

### **Ethical approval**

Health Data is one of the sensitive datasets; therefore, it was collected from the DHIS2 platform with approval provided by Tanzania's Ministry of Health.

**Data Preparations**

The HIV data for Tanzania’s mainland 26 regions recorded from January 2015 to October 2022 were available. For each region, the data were collected in three categories: monthly total tested cases, total positive cases (new and return), and newly tested positive cases. Total tested cases

involve all cases without caring whether they are repeated or new. The total positive case is the total number of positive cases, including repeated cases. The newly tested positive case only includes new cases from tests and excludes the repeated cases. Sample Data for one of the regions is presented in Table 2.

**Table 2:** Sample Dataset from open Tanzania HMIS portal for Arusha Region

Period	Total tested cases (New & Return)	Total positive cases (New & Return)	New positive cases (new cases)
January 2022	20082	414	276
February 2022	18443	430	197
March 2022	21995	424	272

To get estimated sample data for Tanzania, the sum of values for each category was taken. For example, the total tested cases in Tanzania are the sum of the total tested cases for all regions.

In many machine learning models, alignment and correlation of training data reduce prediction error and improve performance. Therefore, it was necessary to normalize the data in each range to minimize data differences. Different input data usually have different dimensions and units. Data normalization leads the models to faster convergence and minimizes errors during training. Then, data splitting followed, whereby the available sequential data set was divided into training, validation and testing sets. Deep learning models were used to predict HIV prevalence in Tanzania. The models were selected based on popularity and their good performances for time series predictions. In this study, RNN, LSTM, GRU and Conv2D were used. Several performance criteria, such as RMSE, MAE and MAPE, were used to compare the performances of the models. Based on the results, the best model proposed for predicting HIV prevalence was obtained.

**Proposed Research Design**

Figure 2 presents the proposed study design for evaluating the performance of deep learning models in predicting the HIV prevalence in Tanzania. The available data were collected from the DHIS2 health information system. The data features monthly tested cases, positive cases, and new cases for all regions in Tanzania. The prediction was considered based on univariate and multivariate data. In univariate prediction, only one feature was considered, and multiple features were considered in multivariate. In univariate predictions, data were converted to time series using rolling window techniques presented in D. A sequential dataset containing the input and corresponding output was obtained using the rolling window techniques.

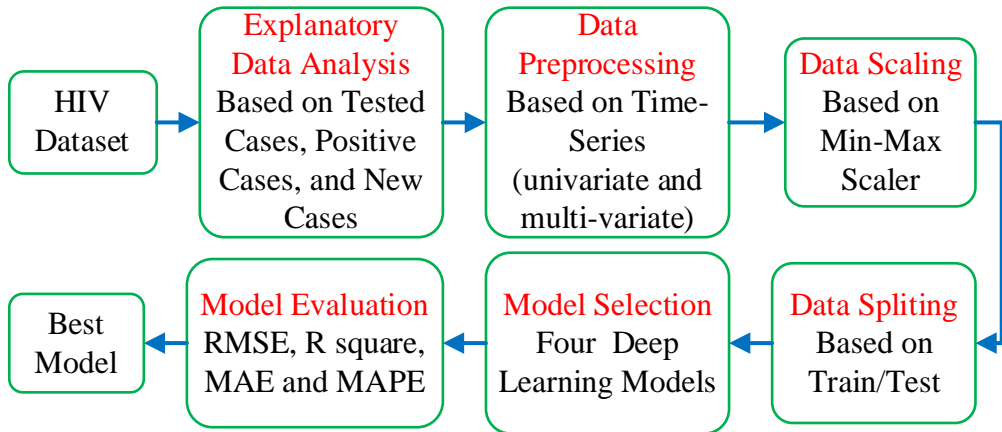


Figure:2 Proposed Research Design for HIV Prediction

**Rolling Window Techniques and Window Sizes**

In machine learning, time series data are the events associated with the time when the events occurred. In deep learning, which is supervised machine learning, the machines must be supplied with data and its corresponding output. Therefore, historical data of the events must be processed to generate datasets of inputs and expected outputs. The data samples depend on how far we should go back to define one sample. In generating such time series data for machine learning, the rolling/sliding window techniques are mostly used, as presented in Figure 3. The two major considerations in these techniques are the size of the window, which is the steps back from the present that

should be used to predict the future. Another consideration is the step ahead that the present data should be used to predict. For example, in Figure 3, the window size of 2 means the input data of the first sample will be (a, b), and the output will be c. Then, the window slides to get the next sample. The number of steps for prediction determines how far forward the model should be able to predict. For example, in Figure 3, consider window size 3, whereby the three-step ahead prediction has been demonstrated. For the input (a, b, c), the output is f, not d. Therefore, the window size and time steps are very important data parameters in determining the performance of the models.

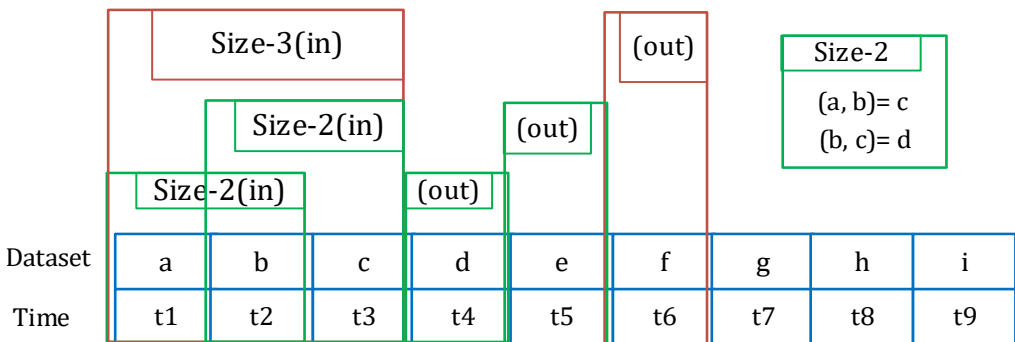


Figure: 3The Sliding/Rolling Window Techniques



### Prediction Datasets Features

The prediction models can be conducted considering the univariate and multivariate data. Since this study focuses on predicting HIV prevalence considering cases, for univariate, the prediction uses the new cases only in the datasets to predict the future. Therefore, in this study, experiments were conducted considering univariate cases. However, considering other features in the dataset highly influences the predicted results but is out of the scope of this work.

### Experimental setup

In this study, two types of experiments were conducted. The first experiment compared the performance of the considered models for different sizes of input datasets, known as window size. The main goal was to check the accuracy of the models for various window sizes. Therefore, experiments were conducted considering different window sizes, such as 2, 4, 6, and 8. The second experiment compared the performance of the considered models in predicting the future. The aim was to investigate how far forward can the model predict accurately using the present history. Therefore, experiments were conducted considering different time steps ahead, such as 1, 2, 4, 6, 8, and 10 months. In each time step, average execution time per epoch and statistical parameters such as MAE, MSE, and MAPE were considered.

### Results and Discussion

#### Parameters Settings

Experiments were conducted using the univariate data sets for 1-step ahead prediction and then for  $l$ -step ahead predictions. Different window sizes, such as 2, 4, 6, 8, and 7, were considered. The Panda Min-Max Scaler tool was used to scale the data between 0 and 1. The data was converted to the sequential dataset using the rolling window technique. The data was divided into training and testing sets at a ratio of 75% and 25%, respectively. According to studies, the choice of size between the training and testing dataset ratio is training dataset is anything between 70% and 80%. In comparison, testing can range from 30% to 20% of the total dataset (Raschka, 2018). Therefore, for this case, training and testing data ratios were arbitrarily chosen depending on the type and size of the data but, referring to the ML guidelines. The models were designed using stacked techniques. In all cases, the parameters of the models are presented in Table 3. The choice of parameters depends on the recommendation from the literature regarding the deep learning models (Sundermeyer et al. 2012). The models were designed using stacked techniques. In all cases, the parameters of the models are presented in Table 3.

**Table 3:** Models Parameters

S/N	RNN	LSTM	GRU	Conv2D
Layers 3 (sizes)	(124, 100,100)	(124, 100,100)	(124, 100,100)	(124, 100,100)
Activation functions	relu	relu	relu	relu
Epochs (max)	1000	1000	1000	1000
Runs	30	30	30	30

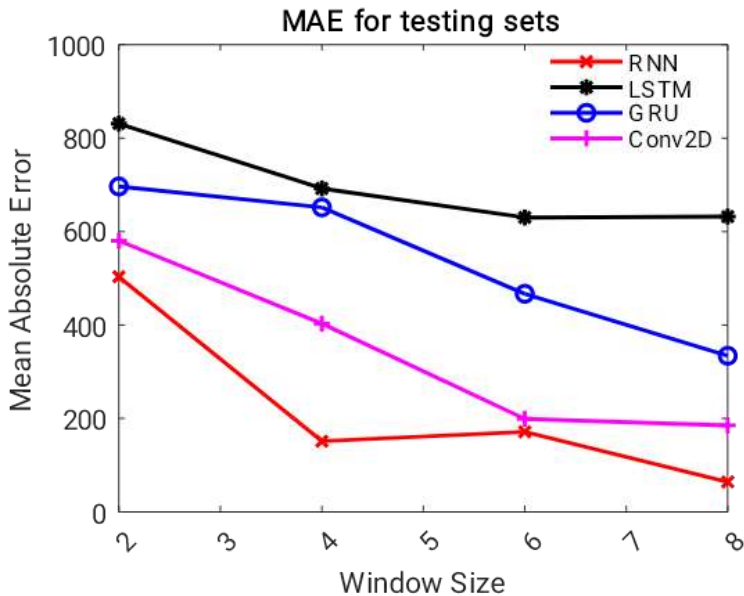
### Results of the influence of window size on model's performance

The results of the experiments for testing the model's accuracy for different window sizes are presented in Table 3. All experiments were run 30 times using 1000 iterations for each run.

Figure 4 presents the results of the MAE experiments that were conducted. The results show that the MAE for all models decreases as the window size increases. The results mean that it becomes easy for the model to accurately predict the future as the size of historical data increases. However, there is no

generic window size for all the models. The RNN has displayed superior results for all window sizes, followed by Conv2D and GRU. Results for other accuracy parameters, such as MSE and MAPE, are presented in Table 4 and Table 5. Results in Table 4 and Table 5 show that the RNN achieved the

lowest MSE and MAPE for all window sizes except window 6, where GRU was the best performer, and window 10, where the LSTM was the best. Considering the sum of MSEs and MAPEs for all cases, the RNN is the best followed by GRU.



**Figure 4:** Mean Absolute Errors per Window

Figure 5 presents the results for the average training time for the models. The results show that Conv2D has displayed superior results for all window sizes, followed by RNN and GRU. Figure 6 presents the prediction results of the algorithms and the actual data. Results show that all the algorithms could follow the trend of the actual data sample.

Results of the influence of window size on the performance of the models imply that, although Conv2D has achieved the best average training time for all window sizes, its accuracy cannot surpass that of RNN and GRU. From this result, the RNN can be considered the best model for the considered window sizes because it has shown the best accuracy and relatively less training time.

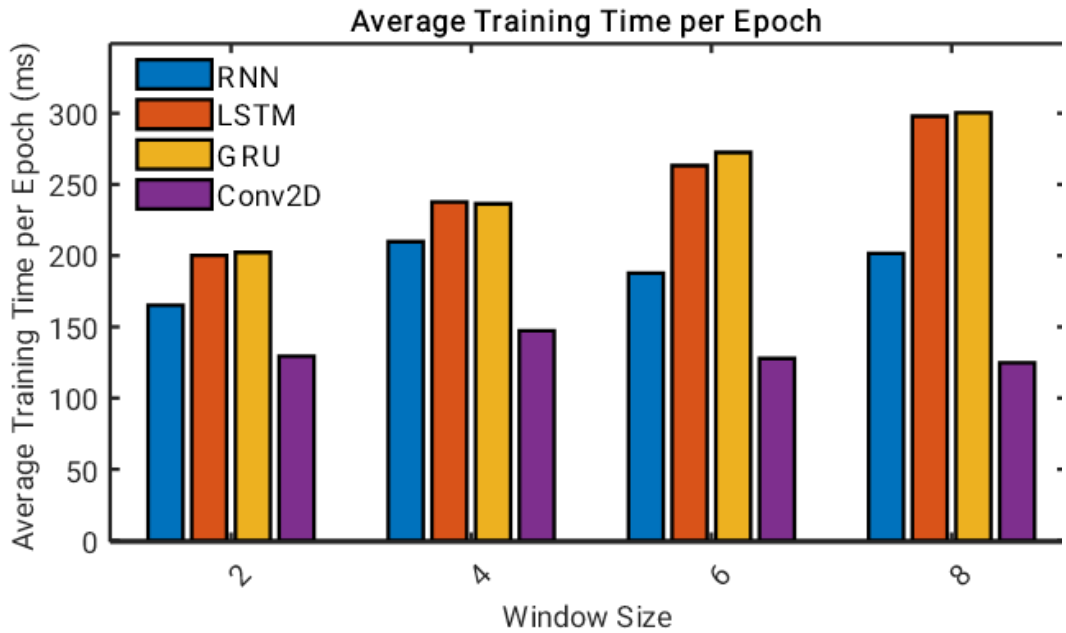


Figure 5: Models Average Training Time

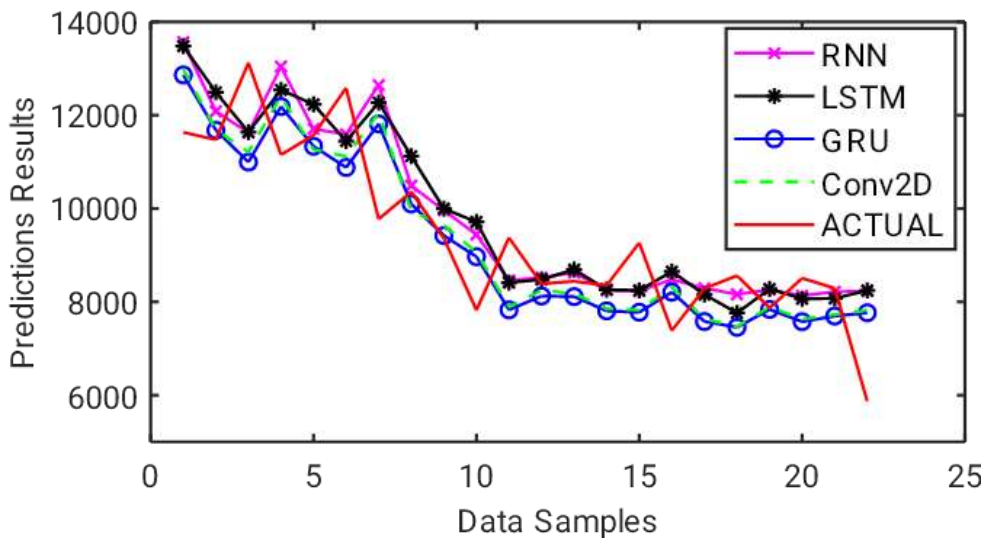


Figure 6: Prediction Results

Table 4: The MSE for Different Window Size

Window Size	RNN	LSTM	GRU	Conv2D
1	3442848.35	14746163	8234570.8	7904328.55
2	6955274.895	20070297.11	16586224.11	15305474.63
4	14106235.71	32319817.65	14388695.35	27961950.76
6	19499278.93	38863214.07	16921084.53	64159262.4
8	74074613.62	84797241.23	75709063.08	79880177.62

10	100527480.4	89785941.27	90997724.45	97058754.36
Total	218605731.9	280582674.3	222837362.3	292269948.3

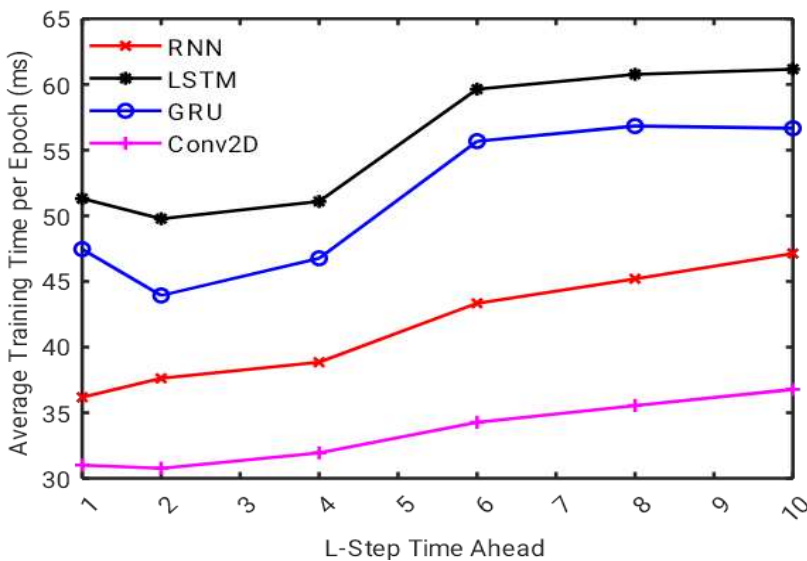
**Table 5:** The MAPE for Different Window Size

Window Size	RNN	LSTM	GRU	Conv2D
1	0.14431715	0.289718979	0.229134607	0.22211818
2	0.2085979	0.32901849	0.30785088	0.29972253
4	0.27414295	0.390928257	0.288260651	0.37082983
6	0.27623533	0.420749923	0.249715644	0.48726508
8	0.50739621	0.528659297	0.514242302	0.52108293
10	0.55284518	0.538400369	0.54012281	0.54819872
Total	1.96353472	2.497475315	2.129326894	2.44921727

**Results of the influence of time steps on the model's performance**

This section presents the results for testing the models' performance for different time steps ahead. All experiments were run 30 times using a window size of 4 and 1000

iterations for each run. Figure 7 presents the results for the average training time for the models. The results show that Conv2D was the fastest, followed by RNN and GRU for all time steps. The LSTM was the slowest.



**Figure 7:** Average Training Time per Epoch

The results in Figure 8 show that as the time steps increase, MAE for all models increases. The results mean that as the time step increases the prediction ability of the models decreases. This result implies that it becomes difficult for the model to predict very far from the present accurately. For the short time steps less than 4, the Conv2D

shows the best performance with a minimal MAE value, followed by the RNN and GRU, while the LSTM shows the worst performance with the largest MAE. As the time steps increase, the accuracy of RNN and GRU relative to other models decreases. Starting from time step 8, the LSTM shows the best performances, followed by GRU and

RNN. The Conv2d shows the worst performance starting from step 6 onward. Results in Table 6 and Table 7 show that the RNN achieved the lowest MSE and MAPE for all time steps except for 6, where GRU was the best performer, and 10 steps, where the LSTM was the best. Considering the sum of MSEs and MAPEs for all steps, the RNN is the best followed by GRU.

Results of the influence of time steps on the performance of the models imply that the performance of the models is affected by the

time steps. The results show that although the average training time for Conv2D is minimal, its accuracy was surpassed by the accuracy of RNN and GRU for all time steps. The results show that, although the LSTM has the largest average training time, its accuracy is the best in the large time step ahead, which proves the superiority of the LSTM in dealing with long-term dependencies.

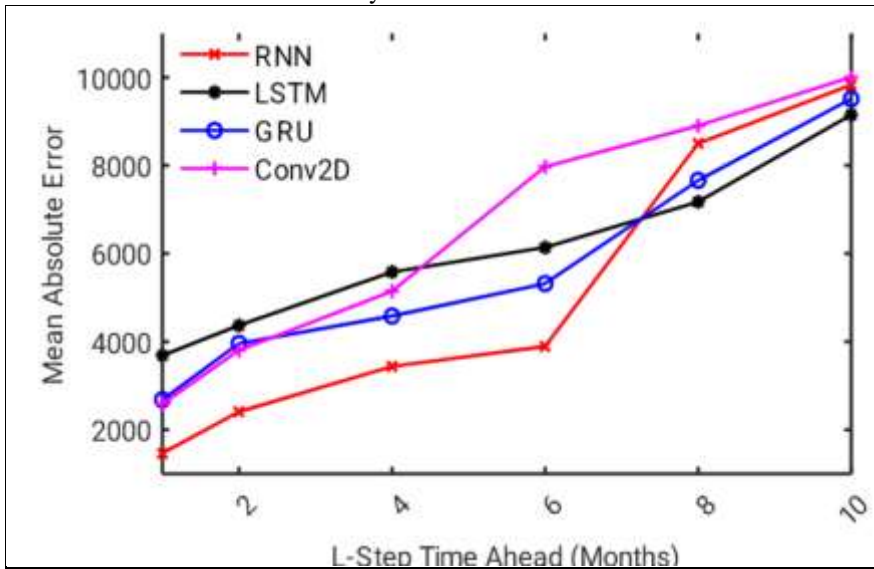


Figure 8: Mean Absolute Errors for Models

Table 6: The MSE for Different Time Steps

Time Steps	RNN	LSTM	GRU	Conv2D
1	3442848.35	14746163	8234570.8	7904328.55
2	6955274.895	20070297.11	16586224.11	15305474.63
4	14106235.71	32319817.65	14388695.35	27961950.76
6	19499278.93	38863214.07	16921084.53	64159262.4
8	74074613.62	84797241.23	75709063.08	79880177.62
10	100527480.4	89785941.27	90997724.45	97058754.36
Total	218605731.9	280582674.3	222837362.3	292269948.3

Table 7: The MAPE for Different Time Steps

Time Steps	RNN	LSTM	GRU	Conv2D
1	0.14431715	0.289718979	0.229134607	0.22211818
2	0.2085979	0.32901849	0.30785088	0.29972253
4	0.27414295	0.390928257	0.288260651	0.37082983

6	0.27623533	0.420749923	0.249715644	0.48726508
8	0.50739621	0.528659297	0.514242302	0.52108293
10	0.55284518	0.538400369	0.54012281	0.54819872
Total	1.41068954	1.959074946	1.589204084	1.90101855

### Conclusion and Future Work

This study has compared the performance of deep learning models in forecasting HIV incidences in Tanzania. The HIV incidence data from twenty-six regions in Tanzania collected from the DHIS2 was considered. The study focused on the influence of the data window size and time steps ahead. Four deep learning models were considered: Conv2D, RNN, GRU and LSTM. The results show that the Conv2D is simple and fast in training, but its accuracy is not the best. For long-term prediction of HIV incidence, the LSTM and GRU show the best performances. The findings of this study show that, in applications where accuracy is critical, the LSTM can be used, and when time is critical, the GRU can be the best candidate.

Despite the excellent performance of these deep learning models, especially the LSTM, in predicting HIV cases, there were still some deviations between the actual and the expected results. Therefore, the accuracy of the models in predicting HIV incidence in Tanzania can be improved by designing more advanced or hybrid deep learning models. Recommended advancements can include combining more than one model or adding other mathematical computational techniques such as empirical model decomposition and metaheuristic algorithm.

### Ethical approval

Health Data is one of the sensitive datasets; therefore, the data was collected from the DHIS2 platform through special approval provided by Tanzania's Ministry of Health.

### Acknowledgement

The University of Dar es salaam and the HISP project at the University of Oslo are highly acknowledged for sponsoring this study.

### Declaration of Interest Statement

The authors have no potential conflict of interest.

Ahlström MG, Ronit A, Omland LH, Vedel S and Obel N 2019 Algorithmic prediction of HIV status using nation-wide electronic registry data. *EclinicalMedicine*. 17:100203.

Alehegn M 2022 Application of machine learning and deep learning for the prediction of HIV/AIDS. *HIV & AIDS Review*. 21(1): 17-23.

Aribe S, Gerardo B and Medina R 2022 Neural Network-based Time Series Forecasting of HIV Epidemics: The Impact of Antiretroviral Therapies in the Philippines. *J. Posit. Psychol*. 6(5): 7844–7855.

Arun Kumar K, Kalaga DV, Kumar CMS, Kawaji M and Brenza T 2022 Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alex. Eng. J*. 61(10): 7585-7603.

Baccouche A, Garcia-Zapirain B, Castillo Olea C and Elmaghraby A 2020 Ensemble deep learning models for heart disease classification: A case study from Mexico. *J. Inf*. 11(4): 207.

Balzer LB, Havlir DV, Kanya MR, Chamie G, Charlebois ED, Clark TD, Koss CA, Kwarisiima D, Ayieko J and Sang N 2020 Machine learning to identify persons at high-risk of human immunodeficiency virus acquisition in rural Kenya and Uganda. *Clin. Infect. Dis*. 71(9): 2326-2333.

Baytas IM, Xiao C, Zhang X, Wang F, Jain AK and Zhou J (2017). Patient subtyping via time-aware LSTM networks.

- Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, Halifax NS Canada.*
- CDC 2019 90-90-90 progress report. Retrieved 7th November, 2023 from <https://www.cdc.gov/hiv/pdf/policies/progresreports/cdc-hiv-preventionprogressreport.pdf>
- Ceia-Hasse A, Sousa CA, Gouveia BR and Capinha C 2023 Forecasting the abundance of disease vectors with deep learning. *Ecol. Inform.* 78: 102272.
- Chen Q, Wen D, Li X, Chen D, Lv H, Zhang J and Gao P 2019 Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow. *PLoS One.* 15(3):e0231199.
- Chikusi HL and Kaijage S 2022 *Machine learning model for prediction and visualization of HIV index testing in northern Tanzania.* Masters thesis and dissertations, The Nelson Mandela African Institution of Science and Technology.
- Dansana D, Kumar R, Das Adhikari J, Mohapatra M, Sharma R, Priyadarshini I and Le D-N 2020 Global forecasting confirmed and fatal cases of COVID-19 outbreak using autoregressive integrated moving average model. *Public Health Front.* 8: 580327.
- DHIS2 (2023) *District health information system 2.* DHIS 2. Retrieved October from <https://dhis2.org/>
- Elsaraiti M and Merabet A 2021 A comparative analysis of the ARIMA and LSTM predictive models and their effectiveness for predicting wind speed. *Energy J.* 14(20): 6782.
- Gu J, Liang L, Song H, Kong Y, Ma R, Hou Y, Zhao J, Liu J, He N and Zhang YJSr 2019 A method for hand-foot-mouth disease prediction using GeoDetector and LSTM model in Guangxi, China. *Sci. Rep.* 9(1): 1-10.
- Hochreiter S and Schmidhuber J 1997 Long short-term memory. *Neural Comput.* 9(8): 1735-1780.
- Jaseena K and Kovoor BC 2021 Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks, *Energy Convers. Manag.* 234: 113944.
- Jasmine PP, Kotecha K, Rajini G, Hariharan K, Utkarsh Raj K, Bhargav Ram K, Indragandhi V, Subramaniaswamy V and Pandya S 2023 Lung diseases detection using various deep learning algorithms. *J. Healthc. Eng.* 023(1): 3563696.
- Jiang W, Liu S, Zhang H, Sun X, Wang S-H, Zhao J and Yan J 2022 CNNG: A convolutional neural networks with gated recurrent units for autism Spectrum disorder classification. *Front. Aging Neurosci.* 14: 948704.
- Kagaayi J, Chang LW, Ssempijja V, Grabowski MK, Ssekubugu R, Nakigozi G, Kigozi G, Serwadda DM, Gray RH and Nalugoda F 2019 Impact of combination HIV interventions on HIV incidence in hyperendemic fishing communities in Uganda: a prospective cohort study. *Lancet HIV.* 6(10): E680-E687.
- Lekey-Kawo G 2023 History, national response, and challenges of HIV in Tanzania. *Orap J.* 4(1): 1002.
- Li X, Xu X, Wang J, Li J, Qin S and Yuan J 2020 Study on prediction model of HIV incidence based on GRU neural network optimized by MHP SO. *IEEE Access.* 8: 49574-49583.
- Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ and Volk JE 2019 Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. *Lancet HIV.* 6(10): E688-E695.
- Mutai CK, McSharry PE, Ngaruye I and Musabanganji EJBmrm 2021 Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa. *BMC Med. Res. Methodol.* 21(1): 1-11.
- Nosratabadi S, Mosavi A, Duan P, Ghamisi P, Filip F, Band SS, Reuter U, Gama J and Gandomi AH 2020 Data science in

- economics: comprehensive review of advanced machine learning and deep learning methods. *J. Math.* 8(10): 1799.
- Nyoni SP and Nyoni T 2021a Forecasting ART coverage in Kenya using the Multilayer perceptron neural network. *Int.J.Innov. Res. Sci. Eng. Technol.* 5(3): 161- 165.
- Nyoni SP and Nyoni T 2021b Rwanda's Art Program Success Story: Insights from the Artificial Neural Networks. *Int.J.Innov. Res. Sci. Eng. Technol.* 5(3): 217-221.
- Obaid OI, Mohammed MA and Mostafa SA 2020 Long short-term memory approach for Coronavirus disease prediction. *Inf. Technol. Manage.* 12: 11-21.
- Raschka S 2018 Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *arXiv preprint.* 1811:12808.
- Sarker IH 2021 Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* 2(3): 1-21.
- Shi H and Zhang S 2022 Accurate prediction of anti-hypertensive peptides based on convolutional neural network and gated recurrent unit. *Interdiscip. Sci.* 14:879–894.
- Stavelin AK 2021 A Mathematical Model for COVID-19 to Predict Daily Cases using Time Series Auto Regressive Integrated Moving Average (ARIMA) Model in Delhi Region, India. *Int.J.Innov. Res. Sci. Eng. Technol.* 6(8): 5.
- Sundermeyer M, Schlüter R and Ney H (2012, September 9-13.). LSTM neural networks for language modeling. *Interspeech, Portland, OR, USA.*
- UNAIDS (2020) *Country progress report - United Republic of Tanzania* (Global AIDS Monitoring 2020. Retrieved 10th May, 2022 from [https://www.unaids.org/sites/default/files/country/documents/TZA\\_2020\\_countryreport.pdf](https://www.unaids.org/sites/default/files/country/documents/TZA_2020_countryreport.pdf).
- UNAIDS (2022) *Global HIV & AIDS statistics-Fact sheet.* Retrieved 21st October, 2022 from [https://www.unaids.org/sites/default/files/media\\_asset/UNAIDS\\_FactSheet\\_en.pdf](https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf).
- Van Houdt G, Mosquera C and Nápoles GJAIR 2020 A review on the long short-term memory model. *Artif. Intell. Rev.* 53(8): 5929-5955.
- Wang G, Wei W, Jiang J, Ning C, Chen H, Huang J, Liang B, Zang N, Liao Y and Chen R 2019a Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. *Epidemiol Infect.* 147:e194.
- Wang H, Li M and Yue X 2021 IncLSTM: incremental ensemble LSTM model towards time series data. *Comput. Electr. Eng.* 92:107156.
- Wang T, Tian Y and Qiu RG 2019b Long short-term memory recurrent neural networks for multiple diseases risk prediction by leveraging longitudinal medical records. *IEEE J. Biomed. Health Inform.* 24(8): 2337-2346.
- Wang Y, Liu L and Wang C 2023 Trends in using deep learning algorithms in biomedical prediction systems. *Fnins.* 17:1256351.
- WHO (2022) *Key facts and latest estimates on the global HIV epidemic.* World Health Organisation. Retrieved 21st October, 2022 from [https://cdn.who.int/media/docs/default-source/hq-hiv-hepatitis-and-stis-library/key-facts-hiv-2021-26july2022.pdf?sfvrsn=8f4e7c93\\_5](https://cdn.who.int/media/docs/default-source/hq-hiv-hepatitis-and-stis-library/key-facts-hiv-2021-26july2022.pdf?sfvrsn=8f4e7c93_5)
- Xia J, Pan S, Zhu M, Cai G, Yan M, Su Q, Yan J and Ning G 2019 A long short-term memory ensemble approach for improving the outcome prediction in intensive care unit. *Comput. Math. Methods Med.* 2019: 8152713.
- Yamak PT, Yujian L and Gadosey PK (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence,* New York; NY; United States.
- Yang E, Zhang H, Guo X, Zang Z, Liu Z and Liu Y 2022 A multivariate multi-step LSTM forecasting model for tuberculosis incidence with model explanation in



- Liaoning Province, China. *BMC Infect. Dis.* 22(1):1-13.
- Yu Y, Si X, Hu C and Zhang J 2019 A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* 31(7): 1235-1270.
- Zhang R, Song H, Chen Q, Wang Y, Wang S and Li Y 2022 Comparison of ARIMA and LSTM for prediction of hemorrhagic fever at different time scales in China. *PLoS One.* 17(1): E0262009.