TRACKING DYNAMICS CHANGE PARAMETERS OF INFECTIOUS DISEASE OUTBREAK WITH BIFURCATED TIME-SERIES LSTM

Adegboyega Adebayo Olumide O. Obe, Francis Osang

Introduction

If the spread of highly infectious diseases is undetected early enough, they can peak at a critical fatality rate in a short period, due to the prevalence of unhygienic social behaviour, and ineffective healthcare policies.

The chaotic phenomenon of outbreaks makes it a perturbed problem both to predict and to combat (Obe *et.al*.,2022).

Interestingly, the deep learning-based time-series predictive modelling of infectious disease outbreaks has gained significant traction.

The difficulty mostly encountered is the inability to represent the causality mechanics which propagate infectious disease outbreaks in deep learning models.

Therefore, theory-based mechanistic methods have become a mainstay of epidemic forecasting due to their ability to capture the underlying causal processes through mathematical and computational representations (Wang *et.al.,*2022).

Research Motivation

This study seeks to validate the prospect of enhancing deep learning algorithms with bifurcation theory (Obe *et.al*.,2022) using pre-vaccination era measles outbreak epidemiological data

Most epidemiological modelling techniques have struggled to quantify uncertainties (Keshavamurthy *et.al.,*2022) that marred the predictions of infectious disease outbreaks (Scarpino *et.al.,* 2019).

[20] assumptions about the prospect of combining techniques from other fields with machine learning could aid to obtain plausible results in the prediction of trends of infectious disease outbreaks.

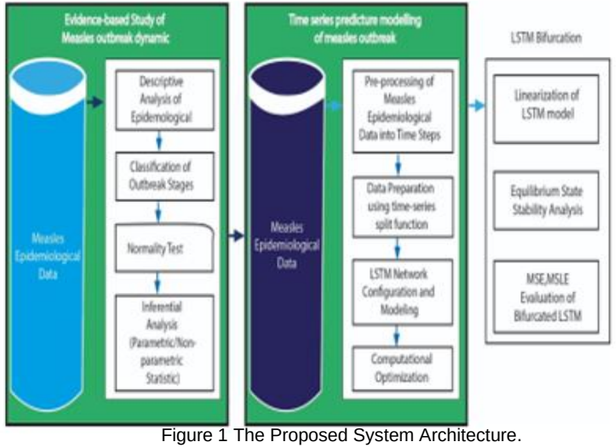

Figure 1 The Proposed System Architecture.

Experimental Dataset

The measles epidemiological data used in this study is public domain license data obtained from Kaggle open-source repository.

The dataset originally belongs to HealthData.gov - Project Tycho(https://healthdata.gov/dataset/Project- Tycho-Level-1-Data/g89t-x93h).

The Infectious Disease Ontology (IDO core) http://purl.obolibrary.org/obo/ido.owl) helps to identify California as the epicentre with recurring waves of the outbreaks were at the average incidence rate of 8.86 per 100,000 population sample.

Exploratory Analysis of Epidemiological Data

Table 1 Summarizes the descriptive analysis of the measles outbreak in California (1928-1967).

Statistical Attribute	Estimate
Epidemic weeks Count	1,457
Average New Cases	833
Maximum weekly cases	6930
75th percentile of cases	¡1229
Average Standard Deviation	1018
Cumulative cases	1213912

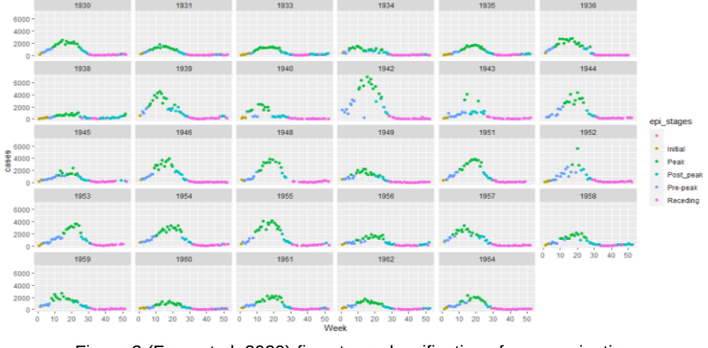

Figure 2 (Feng et.al.,2020) five stage classification of pre-vaccination era measles epidemics in California (1930-1964)

Table 2 Results of quarterly correlation of measles epidemic stages for twenty-five epidemic years in California.

Season	Epidemic stage	Correlation	Statistical Significance
First quarter	Initial, Pre-peak,Peak	0.7092	p-value ¡ 0.05
Second quarter	Peak	-0.4623	p-value ¡ 0.05
Third quarter	Receding	-0.6775	p-value ¡ 0.05
Fourth quarter	Receding	0.3276	p-value ¡ 0.05

LSTM Time Series Predictive Modelling

Table 3: Shows five data samples where four weeks input time steps and one week output time step

Input	Output
[178 442 490 628]	[864]
[442 490 628 864]	[943]
[490 628 864 943]	[954]
[628 864 943 954]	[1151]
[864 943 954 1151]	[1433]

Table 4 Shows the Measles Epidemic data components used to develop the LSTM model.

Data set	Year
Training Data	1930-1956
Validation Data	1962-1957
Test Data	1964

$$\boldsymbol{y(t) = g(\,x(t))} \qquad (1)$$

Where
y(t) : actual output measles cases per week,
x(t) : Input series of Measles cases in a time step ,
g **:** Rn× R*m* →R*n* , static LSTM nonlinear map function.

$$\boldsymbol{y(t) = g(x(t)) \equiv \dot{y}(t)} \qquad (2)$$

where $\dot{y}(t)$ is the static nonlinear LSTM model of measles spread

The LSTM model with the minimum loss function was searched using the following categories of optimization techniques to train 3 three LSTM architectures

- Gradient-Based Optimization,
- Evolutionary Optimization,
- Bayesian Optimization,
- and the Hybrid Optimization.

Table 5 Shows the best models of each optimization technique and the corresponding LSTM architecture.

Model	Loss
GS LSTM	0.1090
GS CNN LSTM	0.1372
PSOS LSTM	0.1259
LMA CNN LSTM	0.1348
BS LSTM	0.1142
BS CNN LSTM	0.1268

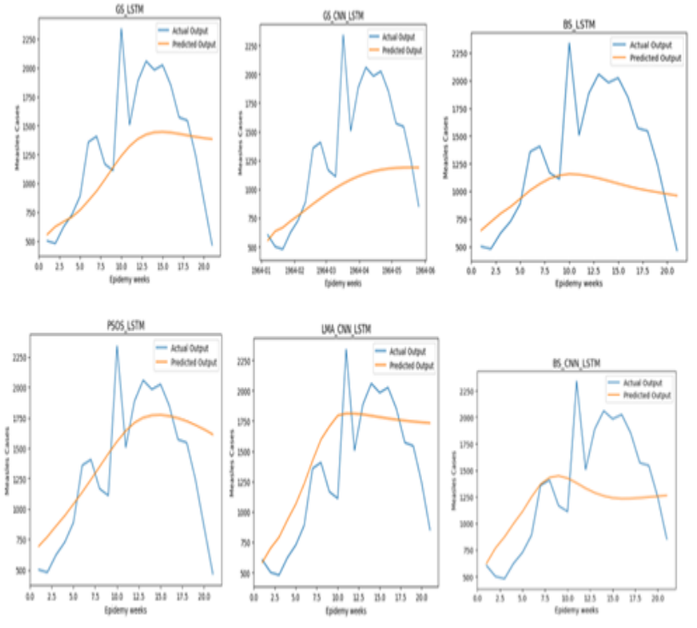

Figure 3 show graphs of the comparing actual and prediction outputs of the prospective regular LSTM models.

Table 6 LSTM models at Increased Input Series.

Model	Loss
GS LSTM	0.0956
BS LSTM	0.0998

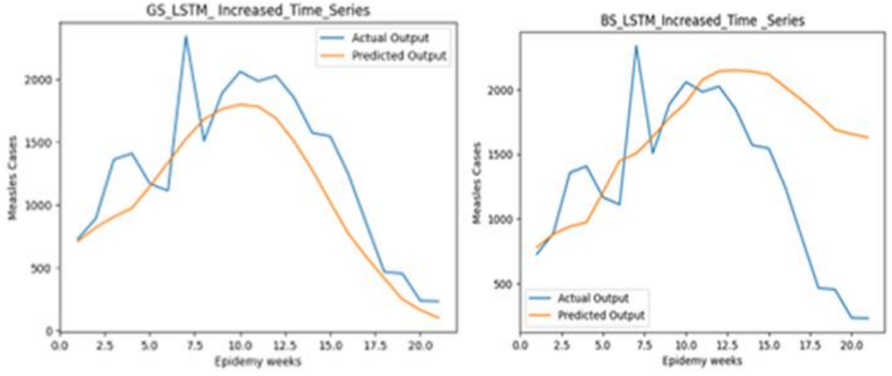

Figure 5 show graphs of the prediction outputs of the prospective regular models at increased input time series.

The architecture and optimization techniques of the LSTM models observed so far have been unable to adequately fit the non-linear relationship of the input data at a low error margin.

Bifurcation Theory

Bifurcation theory involves the search for the trajectory path of changes in model behaviours. To track the trajectory change, the model was normalized in the following dynamic forms:

$$f(x, \mu) = \frac{dx}{dt} = \mu - x^2 \quad (3)$$

$$f(x, \mu) = \frac{dx}{dt} = \mu x - x^2 \quad (4)$$

$$f(x, \mu) = dx = \mu x - x^3 \quad (5)$$

Where μ is the parameter dependency of the dynamic models, and f(x) is the LSTM model to be perturbed.

The task performed was to perturb the high-precision models with the bifurcation canonical forms (equations 3- 5) to a bifurcation point at which the prediction performance of bifurcated LSTM has significantly improved on the regular LSTM

Results

Table 7 Comparing the Evaluations of Best Regular LSTM and Bifurcated-LSTM Models (1962 California Measle Outbreak).

Models	Bifurcation Points	RMSE
GS_LSTM	NA	0.1833
BS_LSTM	NA	0.1892
SADDLE_BIFURCATED_BS_LSTM	1.5	0.1740
PITCHFORK_BIFURCATED_BS_LSTM	1.5	0.1684
TRANSCRITICAL_BIFURCATED_GS_LSTM	1.0	0.1824

Table 8 Comparing the Evaluations of Best Regular LSTM and Bifurcated-LSTM Models (1964 California Measle Outbreak).

Models	Bifurcation Points	RMSE
GS LSTM	NA	0.2962
BS LSTM	NA	0.2933
PITCHFORK BIFURCATED BS LSTM	1.5	0.2776
PITCHFORK BIFURCATED GS LSTM	0.5	0.2907
PITCHFORK BIFURCATED GS LSTM	1.5	0.2882
PITCHFORK BIFURCATED GS LSTM	2.0	0.2916

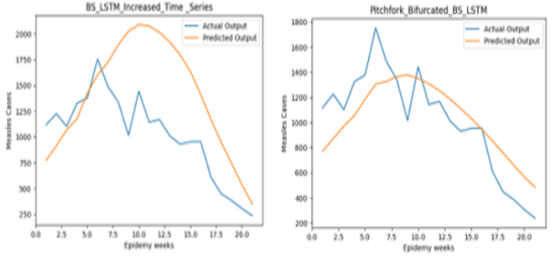

Figure 6 Shows prediction graphs of Bayesian Optimized regular single-layered LSTM and Bayesian trained Pitchfork Bifurcated single-layered LSTM at $\mu = 1.5$ for the first twenty weeks of 1962 California Measles Outbreak.

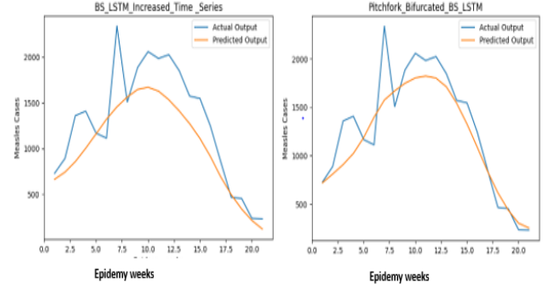

Figure 7 Shows prediction graphs of Bayesian Optimized regular single-layered LSTM and Bayesian trained Pitchfork Bifurcated single-layered LSTM at $\mu = 1.5$ for the first twenty weeks of the 1964 California Measles Outbreak.

Discussion

An optimized measles outbreak LSTM model was enhanced with bifurcation formalisms **Equations 3 − 5** and possible numeric values of control dependency parameters to predict the basic reproduction and bifurcation pattern that represent the changing dynamics of the 1962 and 1964 measles outbreak.

The parameters obtained aided an improved overall prediction performance of LSTM (see Tables 7 & 8).

The pitchfork bifurcated LSTM stimulated a qualitative change in the loss land-scape of the regular LSTM (see Figures 6,7) at the bifurcation point of 1.5 when its' hyperparameters crossed to transit the model from a single stable equilibrium to un-stable which resulted in two additional equilibria [19].

This bifurcation phenomenon influenced the convergence of the model at a new local minimum that generalizes on new data for both the 1962 and 1964 epidemic years, and hence reduced the error margin from the initial stage to the peak stage.

Conclusion

The asymmetric spread of pre-vaccination era measles outbreaks re-occurred at the peak stage of some epidemic years.

The regular LSTM architectures and optimization techniques observed in this study have been unable to learn the non-linear relationship of the epidemiological data at a low error margin.

However, the increase in time steps of the input data led to a trifle reduction of the cost function.

A pitchfork bifurcation of LSTM at bifurcation point =1.5, matches the dynamic change inherent in California's pre-vaccination era measles outbreak at the improved prediction performances of root mean squared error (RMSE) = 0.1684 for the 1962 outbreak, and root mean squared error (RMSE)= 0.2776 for 1964 outbreak.

It could be implied that the parameters of dynamic change of the epidemic was influenced by the basic reproduction of infection agents at $1<R_0<2$.

The dynamics change parameters are dependent on the stochastic nature of the optimization algorithm engaged, and the chaotic nature of an infectious disease outbreak.

References

Obe, O., Sarumi, O. A., Adebayo, A. (2022) Enhancing Epidemiological Surveillance Systems Using Dynamic Modeling: A Scoping Review. Lecture Notes in Networks and Systems, 512–523 https://doi.org/10.1007/978-3-030-96302-6 48

Wang, L., Adiga, A., Chen, J., Lewis, B., Sadilek, A., Venkatramanan, S., Marathe, M. V. (2022) Combining Theory and Data-Driven Approaches for Epidemic Forecasts. Chapman and Hall/CRC eBooks,55–82 https://doi.org/10.1201/9781003143376-3

Keshavamurthy, R., Dixon, S., Pazdernik, K. T., Charles, L. E. (2022) Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches. One Health, 15, 100439. https://doi.org/10.1016/j.onehlt.2022.100439

Scarpino, S. V., Petri, G. (2019). On the predictability of infectious disease outbreaks. Nature Communications, 10(1), —. https://doi.org/10.1038/s41467-019-08616-0

Feng, Z., Xiao, C., Li, P., You, Z., Yin, X., Zheng, F. (2020, December). Comparison of spatio-temporal transmission characteristics of COVID-19 and its mitigation strategies in China and the US. Journal of Geographical Sciences, 30(12), 1963–1984. https://doi.org/10.1007/s11442-020-1822-8