

Forecasting Weekly Influenza Detections in South Africa Using SARIMA and LSTM Models

Ebelechukwu Ajuka¹, Ebenezer Esenogho², Cameron Modisane³

^{1,2,3} Center for Artificial Intelligence and Multidisciplinary Innovation Studies, Department of Auditing, College of Accounting Sciences, University of South Africa, Pretoria, 0002, Gauteng, South Africa

Received:

May 25, 2026

Revised:

July 26, 2026

Accepted:

September 27, 2026

Published:

October 24, 2026

Corresponding Author:

Author Name*:

Ebelechukwu Ajuka

Email*:

ajukaeg@unisa.ac.za

DOI:

10.63158/journalisi.v8i5.1860

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)



Abstract. This study compares Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) models for forecasting weekly laboratory influenza detections in South Africa using World Health Organization (WHO) FluNet surveillance data. Weekly laboratory detection counts from January 1997 to March 2026 were preprocessed, regularized, and treated for missing values, yielding 1,522 weekly observations. Data were divided chronologically into training (20 January 1997–11 May 2020; 1,217 weeks) and a fixed-holdout test period (18 May 2020–16 March 2026; 305 weeks). SARIMA incorporated a 52-week seasonal period, while LSTM used a 52-week lookback window with Min–Max scaling fitted exclusively to training data. A Seasonal Naïve model with a 52-week seasonal lag was included as a benchmark. Forecast performance was assessed using RMSE and MAE, while Diebold–Mariano (DM) test assessed differences in forecast accuracy. LSTM achieved the lowest errors, with RMSE of 9.49 and MAE of 6.31, compared with 23.58 and 15.56 for SARIMA and 29.93 and 16.02 for Seasonal Naïve, respectively. Relative to SARIMA, LSTM reduced RMSE and MAE by 59.8% and 59.4%. DM tests indicated significantly lower forecast loss for LSTM than SARIMA using MSE (DM = 6.6661, $p < 0.001$) and MAE (DM = 9.7569, $p < 0.001$). The study provides an empirical comparison of statistical and deep-learning approaches using a nearly three-decade WHO FluNet surveillance series. LSTM provided significantly more accurate forecasts than SARIMA under the fixed-holdout design. Future research should examine robustness using rolling-origin validation and multiple forecasting horizons.

Keywords: Influenza surveillance, laboratory detections, FluNet, time-series forecasting, SARIMA, LSTM, public health surveillance, South Africa

1. INTRODUCTION

Influenza, commonly known as the flu, is an acute respiratory disease caused by influenza viruses. It is primarily transmitted through airborne droplets released when infected individuals cough or sneeze and can also be transmitted through contact with contaminated surfaces followed by touching the nose, mouth, or eyes. Symptoms, which usually begin within two days after infection, include sudden onset of fever, cough, headache, sore throat, muscle and joint pain, severe malaise, and runny nose [1]. Influenza can also exacerbate underlying chronic conditions and, in severe cases, lead to complications such as pneumonia and sepsis [1]. Due to its high transmissibility and seasonal recurrence, influenza is responsible for an estimated 3–5 million cases of severe illness and approximately 290,000–650,000 respiratory deaths globally each year [1], [2], [3]. The disease also imposes a substantial socioeconomic burden through workplace and school absenteeism, reduced productivity, increased healthcare expenditure, and costs associated with prevention and control measures [4].

Influenza is a serious threat to public health because it causes substantial disease and mortality worldwide. In South Africa, influenza epidemics mainly occur during the winter months, but their timing, duration, magnitude, and prevalent virus strains may vary considerably within and among years. For instance, the 2025 influenza season in South Africa was reported between March and July [5]. Surveillance systems, including the Global Influenza Surveillance and Response System (GISRS), Viral Watch (VW), and Severe Acute Respiratory Illness (SARI), can generate crucial epidemiological data for the effective monitoring of the disease's spread. In addition, the increasing availability of digital and online data has created opportunities for improving disease surveillance and early detection, as demonstrated in studies of influenza-like illness (ILI) [6], [7], [8], [9], monkeypox [10], Zika [11], Ebola [12], [13], chronic kidney disease [14] COVID-19 [15], [16], [17], and tuberculosis [18], [19]. However, conventional surveillance systems may be affected by reporting delays, incomplete reporting, and limited spatial or temporal granularity, which can constrain timely identification of changes in influenza activity. Forecasting models can therefore complement surveillance by providing estimates of future disease activity and supporting preparedness and resource planning.

Recent advances in statistical modelling, machine learning, and artificial intelligence have provided additional approaches for forecasting infectious diseases. Common approaches include autoregressive integrated moving average (ARIMA) models, seasonal ARIMA (SARIMA), Holt-Winters and other exponential-smoothing methods, as well as machine-learning and deep-learning models such as support vector machines, random forests, elastic net, and Long Short-Term Memory (LSTM) networks [4], [5], [6], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. Conventional statistical models provide important benchmarks because they explicitly model temporal dependence, trends, and, where appropriate, recurring seasonal patterns. For example, ARIMA-based approaches have been used to forecast ILI and other infectious diseases, demonstrating the usefulness of historical surveillance observations for anticipating future disease activity [32]. Other comparative studies have shown that the relative performance of forecasting models can vary according to disease, setting, data characteristics, and forecasting horizon. For example, [31] compared ARIMA, GM (1,1), and LSTM models for forecasting tuberculosis cases in mainland China and reported superior predictive accuracy for LSTM, whereas [33] found that a Holt-Winters multiplicative model outperformed ARIMA and SARIMA models for malaria forecasting in Pakistan. These results suggest that model performance should be established by experimentation rather than assuming that either standard statistical or more sophisticated machine-learning approaches will always work best.

With regard to influenza, seasonality is important because disease activity tends to follow a cyclical pattern. SARIMA is therefore a suitable conventional time-series benchmark because it extends ARIMA by explicitly incorporating seasonal autoregressive, differencing, and moving-average components. Its principal advantage is its ability to represent autocorrelation and recurring seasonal structure in a relatively interpretable framework. However, SARIMA is fundamentally based on a specified statistical structure and may have difficulty representing complex nonlinear relationships. LSTM networks provide a complementary modelling approach because they are recurrent neural networks designed to learn dependencies across sequential observations and can capture nonlinear temporal relationships. Their greater flexibility, however, comes at the cost of increased model complexity and the need for careful training and temporal validation. Consequently, comparing SARIMA and LSTM provides an informative test of

whether the additional flexibility of a nonlinear deep-learning model produces meaningful forecasting gains over a conventional seasonal statistical benchmark.

Several studies have demonstrated the application of LSTM and related deep-learning approaches to infectious-disease forecasting. For instance, [34] applied LSTM to vector-borne disease forecasting, particularly West Nile virus, while [35] developed an LSTM-based framework for forecasting COVID-19 trends in Canada and other countries. Although these studies demonstrate the potential of LSTM for disease forecasting, their findings cannot necessarily be generalized to influenza surveillance in South Africa because disease dynamics, surveillance practices, seasonal patterns, and data-generation processes differ across diseases and geographical settings. More importantly, many previous studies have focused on disease incidence, reported cases, or other epidemiological outcomes rather than laboratory-confirmed influenza detections derived from routine surveillance data.

Epidemiological studies have widely applied ARIMA, SARIMA, LSTM, and other machine-learning approaches to forecast disease occurrence and analyze temporal trends [2], [30], [31], [32], [33], [34], [35], [36]. However, relatively few studies have examined influenza forecasting using surveillance data from African countries, particularly South Africa [4], [6], [25], [37]. Previous South African studies have primarily focused on exploring the application of data from the Google quest for forecasting purposes [6], [38], [39], including relationships between search volumes for ILI-related terms and observed ILI detections. While these studies demonstrate the potential value of digital search data for influenza surveillance, they do not directly address forecasting future laboratory influenza detections using historical surveillance observations alone. Furthermore, limited evidence is available on the comparative performance of conventional seasonal statistical and deep-learning approaches for forecasting weekly influenza detections in South Africa. This is a major limitation to this approach, given the strong seasonality of influenza, and characteristics of the surveillance itself.

As such, the current study aims to compare the accuracy of Seasonal Autoregressive Integrated Moving Average and Long Short-Term Memory techniques in forecasting weekly laboratory influenza detections in South Africa using World Health Organization

FluNet surveillance data. Specifically, the research will focus on the following research questions:

- 1) What are the temporal and seasonal patterns of weekly laboratory influenza detections in South Africa according to WHO FluNet surveillance data?
- 2) How accurately can SARIMA and LSTM predict weekly laboratory influenza detections from past surveillance time series?
- 3) Under the study's evaluation design, is LSTM's out-of-sample point-forecast accuracy and prediction-interval performance associated with higher accuracy than that of SARIMA?

To answer the research questions, the study will pursue the following objectives:

1. Describe the temporal and seasonal patterns of weekly laboratory influenza detections in South Africa based on WHO FluNet surveillance data;
- 2) Develop a SARIMA model for predicting weekly laboratory influenza detections based on the time series' temporal and seasonal characteristics;
- 3) Develop a LSTM model for predicting weekly laboratory influenza detections based on the time series' temporal patterns;
- 4) Evaluate and compare the out-of-sample point-forecast accuracy and 95% prediction-interval performance of SARIMA and LSTM using appropriate statistical measures.

The Global Influenza Surveillance and Response System (GISRS) is an established global surveillance network that has been in operation since 1952, with the objective of monitoring influenza viruses and their impact on public health, as well as informing the global influenza response [40]. With the participation of over 100 countries through National Influenza Centers (NICs), WHO Collaborating Centers, and specialized laboratories, GISRS enables the collection, analysis, and sharing of influenza-related data across borders. South Africa is among the countries that participate in the GISRS network, and the nation's influenza surveillance system aligns closely with the goals and objectives of the global surveillance system.

National Institute for Communicable Diseases (NICD) is the designated National Influenza Centre in South Africa under the GISRS framework. As such, it plays a critical role in contributing to the global understanding of influenza by providing epidemiological and

virological information on the disease. In turn, South Africa benefits from the technical cooperation, resources, and expertise offered by the global network. South Africa's influenza surveillance system comprises two main sentinel networks: Surveillance for Viral Watch and Severe Acute Respiratory Illness (SARI). Viral Watch is a surveillance system that monitors influenza-like illness (ILI) in communities as a means of detecting trends in influenza epidemics. Surveillance of SARI, in turn, focuses on the severe forms of acute respiratory infections, including those among hospitalized patients. Together, these networks generate important data that contribute to decision-making at national and international levels in the context of influenza.

Notably, surveillance based in laboratories is an integral component of South Africa's influenza surveillance system, as it underpins the national system's contribution to the global network. Specimens obtained through ILI and SARI surveillance undergo standard diagnostic assays, including polymerase chain reaction (PCR). Influenza viruses isolated from these specimens are characterized in terms of their genetic and antigenic properties, and representative samples are sent to the World Health Organization (WHO) Collaborating Centers as part of the GISRS system. Information sharing and awareness are key aspects of South Africa's participation in the GISRS. Through the FluNet and FluID systems, South Africa contributes to the global influenza surveillance system by virtue of its own data, while also benefiting from the data provided by other participating countries. This enables local health authorities to stay informed about trends in influenza globally as well as to better understand their role in the global context.

Furthermore, GISRS plays a role in capacity-building and technical assistance in South Africa. Through its relationships with the WHO and other stakeholders, South Africa benefits from the training, quality assurance, and research opportunities, as well as the most up-to-date technological innovations in the field of influenza surveillance and virology that are available. GISRS and South Africa's influenza surveillance have a mutually beneficial relationship as South Africa strengthens the capacity of the global network through its contributions of high-quality data and specimens, and the national system benefits from the support of the international network. Specifically, the FluNet system provides essential forecasting and early warning capabilities that are critical to South Africa's influenza surveillance and response activities. In short, the strong

partnership between GISRS and South Africa is crucial to the country's ability to monitor and respond to influenza outbreaks at national and international levels.

FluNet influenza detections refer to laboratory-confirmed influenza cases reported through participating surveillance systems. These detections are interpreted as indicators of influenza activity rather than direct measures of incidence within populations [40]. Detection counts are affected by factors such as surveillance intensity, testing practices, reporting rates, sampling strategies, and changing healthcare-seeking behaviors. These factors require careful consideration when interpreting FluNet data, particularly in the context of the COVID-19 pandemic, which has had a significant impact on both influenza epidemiology and surveillance practices [41], [42].

The Autoregressive Integrated Moving Average (ARIMA) model is a statistical time series forecasting approach that combines autoregressive (AR), differencing (integrated, I), and moving-average (MA) components [36], [43], [44], [45], [46]. An ARIMA model is commonly specified as $ARIMA(p, d, q)$, where p denotes the order of the autoregressive component, d represents the degree of non-seasonal differencing, and q denotes the order of the moving-average component. The autoregressive component models the relationship between the current observation and its previous values, whereas the moving average component represents dependence on previous forecast errors. Differencing is applied, where necessary, to achieve stationarity. The general ARIMA model can be expressed using the backshift operator as:

$$\phi(B)(1 - B)^d y_t = c + \theta(B)\varepsilon_t \quad (1)$$

where y_t represents the observed value of the time series at time t , B is the backshift operator such that $By_t = y_{t-1}$, c is the constant term, and ε_t represents the white-noise error term. The non-seasonal autoregressive polynomial is defined as

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

while the non-seasonal moving-average polynomial is

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q.$$

Here, ϕ_i and θ_j denote the autoregressive and moving-average coefficients, respectively. The differencing operator $(1 - B)^d$ is used to remove non-seasonal non-stationarity when required.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the ARIMA framework by incorporating autoregressive, differencing, and moving-average components at a specified seasonal frequency. A SARIMA model is represented as SARIMA $(p, d, q) \times (P, D, Q)_s$, where P , D , and Q denote the seasonal autoregressive order, seasonal differencing order, and seasonal moving-average order, respectively, and s represents the seasonal period [36], [47], [48], [49]. The general SARIMA model is expressed as:

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = c + \theta(B)\Theta(B^s)\varepsilon_t \quad (2)$$

where $\phi(B)$ and $\theta(B)$ are the non-seasonal autoregressive and moving-average polynomials, respectively, while $\Phi(B^s)$ and $\Theta(B^s)$ represent their seasonal counterparts. The operator $(1 - B)^d$ is the non-seasonal differencing, while $(1 - B^s)^D$ is the seasonal differencing.

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed to learn temporal dependencies in sequential data. LSTM networks were introduced by Hochreiter and Schmidhuber [50] to address limitations associated with conventional recurrent neural networks, particularly vanishing-gradient problems that can make learning long-range dependencies difficult. An LSTM incorporates a memory cell and three principal gating mechanisms: the forget gate, input gate, and output gate which regulate the retention, incorporation, and transmission of information through the network [51].

For an input vector x_t and previous hidden state h_{t-1} , the forget gate determines which information from the previous cell state should be retained:

$$f_t = \sigma(W_f[x_t, h_{t-1}] + b_f) \quad (3)$$

where f_t is the forget-gate activation, W_f is the corresponding weight matrix, b_f is the bias vector, and σ denotes the sigmoid activation function.

The input gate determines which new information is incorporated into the cell state:

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \quad (4)$$

A candidate cell-state value is then generated as:

$$\tilde{C}_t = \tanh(W_C[x_t, h_{t-1}] + b_C) \quad (5)$$

where $\tilde{C}_t$ represents the candidate information that may be incorporated into the memory cell. The cell state is updated according to:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

where C_t represents the updated cell state and $\odot$ notes element-wise multiplication.

The output gate determines which information from the updated cell state is passed to the hidden state:

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o) \quad (7)$$

and the hidden state is subsequently calculated as:

$$h_t = o_t \odot \tanh(C_t). \quad (8)$$

Here, W_i , W_C , and W_o represent weight matrices; b_i , b_C , and b_o represent bias vectors; i_t , o_t , and f_t represent the input, output, and forget gates, respectively; C_t is the cell state; and h_t is the hidden state.

Recent studies have shown that machine learning models outperform traditional models in predicting influenza outbreaks by considering multiple data sources. However, they are vulnerable to outliers and noise in data, which may decrease their prediction precision [52], [53], [54]. Machine learning models, including the LSTM model, are appropriate for complex forecasting tasks because they are able to learn long-range dependencies and capture the underlying nonlinear relationships in the time series data. They can be trained end-to-end without requiring any manual feature engineering and can handle noisy data as well as missing values. However, LSTM models are computationally intensive and typically require large volumes of training data to achieve optimal performance. They are also prone to overfitting when the model becomes overly complex or when the available

training data are insufficient [55], [56], [57]. In this study, SARIMA and LSTM were selected to provide a theoretically complementary comparison between classical statistical and nonlinear deep-learning approaches to forecasting weekly laboratory influenza detections. The selection is particularly relevant given the substantial changes in influenza transmission dynamics and surveillance practices during the COVID-19 pandemic. SARIMA provides an established framework for modelling autocorrelation, differencing, and recurrent seasonal patterns in the influenza series, thereby assessing whether the historical temporal and seasonal structure remain informative under periods of epidemiological and surveillance disruption.

In contrast, LSTM provides a nonlinear sequence-learning framework capable of learning complex temporal dependencies and potentially adapting to changes in the underlying dynamics without requiring an a priori linear functional form. The comparison therefore enables assessment of whether the additional flexibility of nonlinear deep learning provides predictive advantages over a structured statistical time-series model under disrupted surveillance conditions. Other forecasting approaches, including ETS/Holt-Winters, Prophet, TBATS, seasonal naïve, ARIMAX, and hybrid models, were not included in the comparison because the study was designed to evaluate the predictive performance of two complementary modelling paradigms while maintaining a parsimonious and methodologically controlled comparison.

2. MATERIALS AND METHODS

2.1. Study Design and Forecasting Workflow

This study adopted a comparative time series forecasting approach to evaluate the performance of statistical and deep learning models in predicting weekly influenza counts in South Africa. The SARIMA and LSTM models were applied and compared with each other using weekly influenza virological surveillance data. The workflow of this study is shown in Figure 1.

This study adopted a comparative time series forecasting approach to evaluate the performance of statistical and deep learning models in predicting weekly influenza counts in South Africa. Figure 1 below shows the analytical workflow that starts with the collection of the data and preparation. It then proceeds to the development of the

forecasting models, their evaluation, and a direct comparison of the two models. The figure is divided according to the six functional boxes that were used to develop the methodology. These boxes include: data collection; forecasting models; SARIMA model; LSTM model; performance assessment; and, comparison of SARIMA and LSTM.

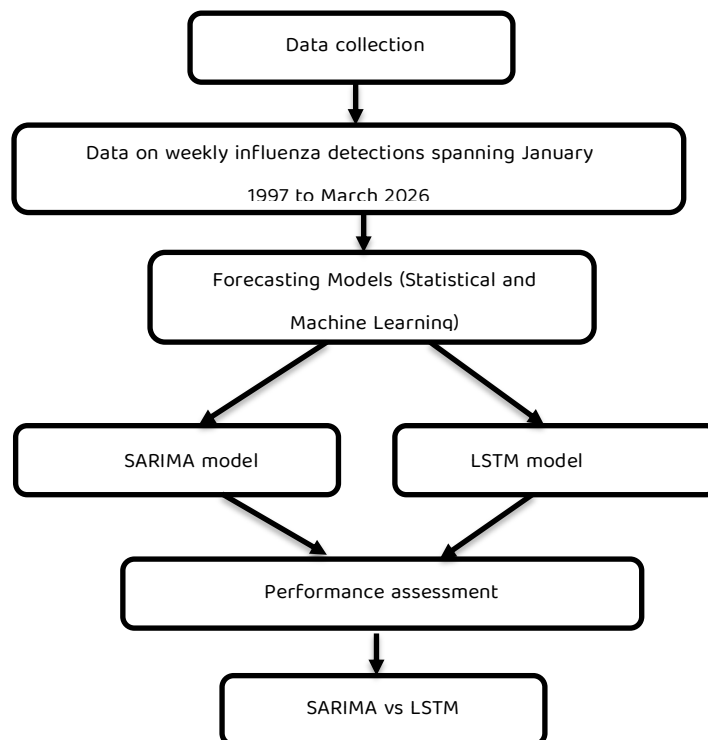


Figure 1. Influenza forecasting workflow for South Africa based on SARIMA and LSTM

The Data Collection box identifies the WHO FluNet surveillance series used as the modelling input. The Forecasting Models box separates the two complementary modelling paradigms evaluated in this study: a statistical SARIMA model and a recurrent deep-learning LSTM model. The SARIMA and LSTM boxes specify the model-specific preparation and estimation steps, including stationarity/seasonal identification for SARIMA and scaling, sequence construction, and recurrent-network training for LSTM. The Performance Assessment box contains the chronological holdout evaluation and the point- and interval-based measures. The final SARIMA versus LSTM box represents the statistical comparison of forecast errors using the Diebold–Mariano test. This organization ensures that each step in the workflow is reproducible and corresponds directly to the analyses reported in Section 3.

2.2. Data Collection

Weekly influenza surveillance data for South Africa were obtained from the World Health Organization FluNet platform on April 1, 2026, covering January 1997 through March 2026 [41]. FluNet is a global virological surveillance resource that receives data from National Influenza Centers, national reference laboratories, and WHO regional databases. The main target variable was INF_ALL, indexed by ISO_WEEKSTARTDATE. The series used in forecasting characterizes laboratory influenza detections and not population-adjusted incidence. Table 1 presents the first five and last five observations of the South Africa FluNet series.

Table 1. First five Observations at the beginning and the last five observations at the end of the South Africa FluNet series

ISO week start	ISO YEAR	ISO WEEK	INF A	INF B	INF ALL
1997-01-20	1997	4	1	-	1
1997-02-17	1997	8	-	2	2
1997-03-10	1997	11	1	-	1
1997-05-12	1997	20	1	-	1
1997-05-19	1997	21	1	-	1
2026-02-16	2026	8	6	4	10
2026-02-23	2026	9	12	2	14
2026-03-02	2026	10	9	2	11
2026-03-09	2026	11	23	3	26
2026-03-16	2026	12	9	3	12

The raw South Africa file contained 1,318 weekly records. Regularization to a complete weekly calendar produced 1,522 weekly observations spanning 20 January 1997 to 16 March 2026. The accompanying surveillance variables, including influenza A, influenza B, specimens processed, and negative specimens, were retained for classification of observed zeros and missing target values.

2.3. Data Preprocessing

The ISO_WEEKSTARTDATE field was converted to datetime format, the series was sorted chronologically, and the target INF_ALL variable was extracted. A complete weekly calendar was created to identify calendar gaps as well as records containing missing

INF_ALL values. Laboratory-supported zeros were distinguished from missing observations using the available specimen-processing and influenza-result fields. When specimens had been processed, INF_A and INF_B were zero, and the number of influenza-negative specimens equaled the number processed, INF_ALL was retained as a true zero. The remaining unavailable values were treated as missing. Table 2 below shows the summary of the dataset after data preprocessing.

Table 2. Data preprocessing summary

Category	Number
Complete weekly calendar	1522
Observed INF_ALL	943
Confirmed zero detections	98
Missing before imputation	481
Kalman-imputed observations	481
Final modelling observations	1522

Of the 1,522 modelling weeks, 943 contained observed INF_ALL values, 98 were classified as laboratory-supported zero detections, and 481 were missing and estimated using state-space Kalman smoothing with a 52-week seasonal component. This approach preserved the temporal dependence and seasonal structure required for SARIMA and LSTM modelling. The final modelling dataset comprised 1,522 consecutive weekly observations with no remaining missing values. ISO Week 53 was retained in the chronological series; however, for annual seasonal summaries, it was mapped to Week 52 so that the seasonal visualization used a standardized 52-week cycle.

2.4. Forecasting Models

Two forecasting paradigms were implemented using the same weekly target series. SARIMA was used as the structured statistical model and LSTM as the recurrent deep-learning model. A Seasonal Naïve forecast based on a 52-week lag was also generated as a benchmark. No exogenous predictors were supplied to either SARIMA or LSTM.

2.4.1. Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA model extends ARIMA by incorporating seasonal autoregressive, differencing, and moving-average components. The model was estimated with the statsmodels SARIMAX implementation using `exog=None`; consequently, the fitted specification is a univariate SARIMA model rather than a model with exogenous regressors. Model identification used the ADF test for non-seasonal stationarity, ACF/PACF inspection for temporal and seasonal dependence, and a 52-week seasonal period. Candidate specifications were compared using AIC and BIC, supported by residual diagnostics and fixed-holdout forecasting performance. The final specification was SARIMA(1,0,1)×(1,1,1)₅₂.

2.4.2. Long Short-Term Memory (LSTM)

The LSTM model used a 52-week lookback window. Min–Max scaling was fitted only to the training observations and then applied to the test observations to avoid information leakage. The network contained two stacked LSTM layers with 64 and 32 units, respectively, each followed by dropout of 0.20, and a one-unit Dense output layer. The network had 29,345 trainable parameters. Adam optimization used a learning rate of 0.001 and MSE loss, with batch size 32 and a maximum of 100 epochs. Ten percent of training sequences were reserved for validation, early stopping used a patience of 10 epochs with restoration of the best weights, and random seed 42 was used for reproducibility, but the training did not shuffle the sequences.

2.4.3. Seasonal Naïve Benchmark

The Seasonal Naïve benchmark forecast the value for each test week from the observation 52 weeks earlier. This provides a transparent seasonal-persistence reference against which the incremental predictive value of SARIMA and LSTM can be assessed.

2.5. Performance Assessment

A fixed chronological holdout was used for out-of-sample evaluation. The full weekly influenza detection series comprises 1,522 observations, spanning from 20 January 1997 to 16 March 2026 (Figure 2a).

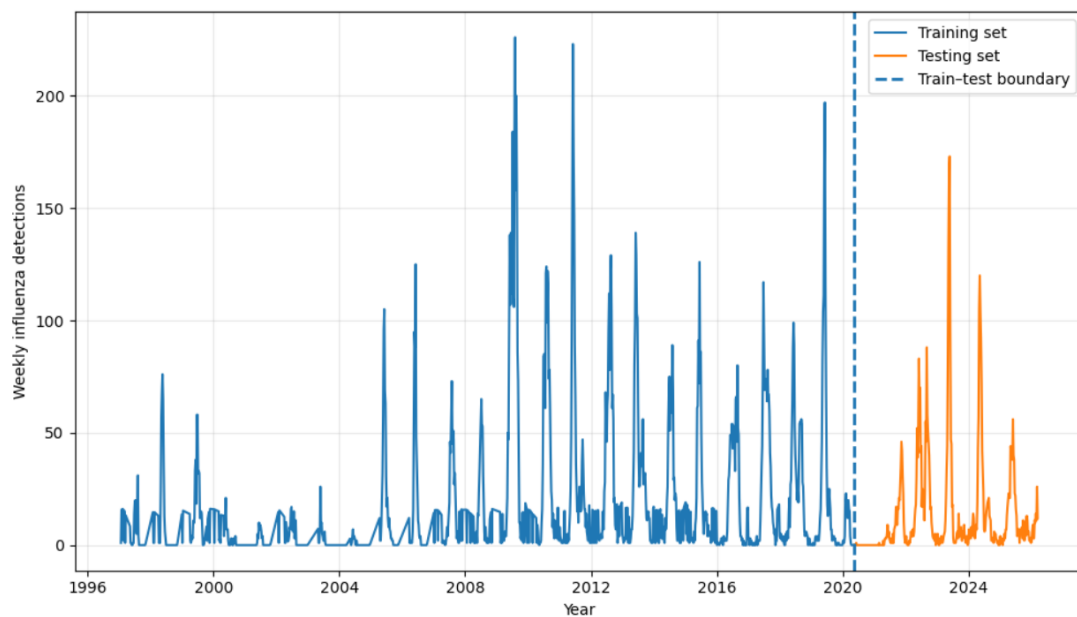


Figure 2a. Chronological training and Testing Periods for Weekly Influenza forecasting

The first 1,217 observations (20 January 1997–11 May 2020) formed the training set and the remaining 305 observations (18 May 2020–16 March 2026) formed the test set (Figure 2b and 2c respectively). No random shuffling was used. For LSTM, the scaler was fitted exclusively on the training period. Model accuracy was evaluated using RMSE and MAE. Predictive uncertainty was assessed using empirical 95% interval coverage and mean prediction-interval width. The uncertainty intervals were model-specific and were not evaluated within a common rolling-origin framework.

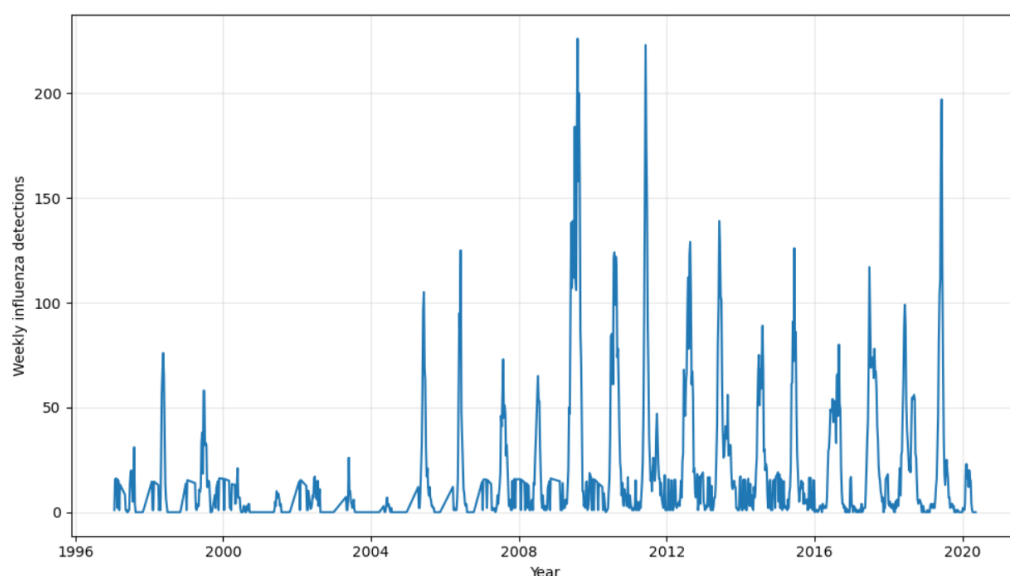


Figure 2b. Chronological Training Periods for Weekly Influenza Forecasting

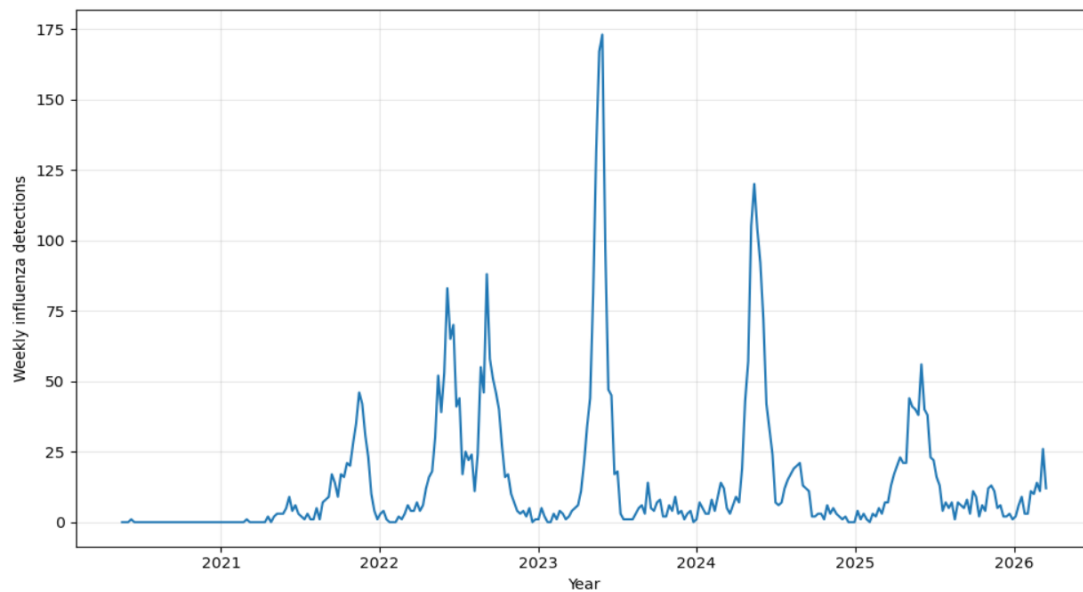


Figure 2c. Chronological Testing Periods for Weekly Influenza Forecasting

2.6. Model Comparison

The Diebold–Mariano (DM) test was used to examine whether the forecast losses from SARIMA and LSTM differed significantly over the fixed test period. The comparison was conducted under squared-error (MSE) and absolute-error (MAE) loss. Positive statistics indicate greater loss for SARIMA because the loss differential was defined as SARIMA loss minus LSTM loss. The DM test complements RMSE and MAE by evaluating whether the observed difference in forecast accuracy is statistically distinguishable under the specified holdout design.

3. RESULTS AND DISCUSSION

3.1. Temporal and Seasonal Characteristics of the Surveillance Series

Figure 3 shows the weekly influenza detections in South Africa including the observed and missing values, Kalman smoothing-based imputation, and the overall temporal pattern of the series. The completed weekly series showed recurrent annual peaks with substantial inter-annual variation in their timing and magnitude. The series also contained low-detection periods between major seasonal peaks and a pronounced disruption in reported influenza activity during the COVID-19 period. The pattern motivated the use of a 52-week seasonal period while also indicating that annual persistence alone may not

capture all short-term changes. The distribution of weekly influenza detections by epidemiological week is shown in Figure 4.

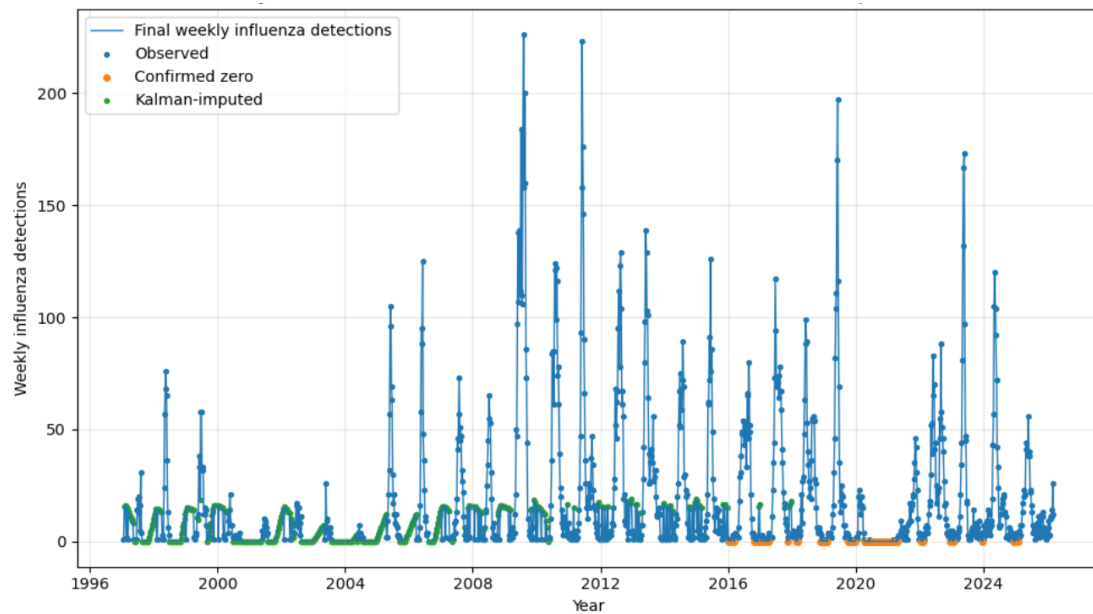


Figure 3. Weekly Influenza Detections in South Africa: Observed and Kalman-Imputed Series, 1997–2026

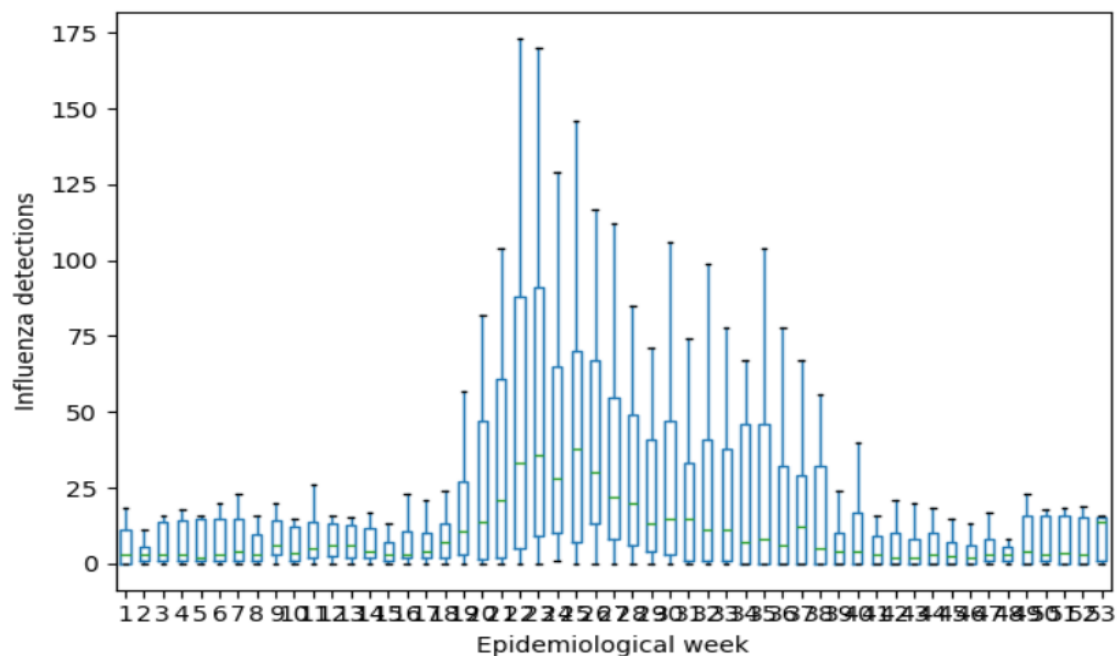


Figure 4. Distribution of Weekly Influenza Detections by Epidemiological Week, 1997–2026

The epidemiological-week distribution in figure 4 showed the concentration of detections during the principal influenza transmission period, with variability in the strength of activity across weeks. The concentration pattern provided empirical support for representing annual seasonality in the forecasting models. Figure 5 shows the annual–Seasonal Heatmap to further demonstrate the seasonal peaks and inter-annual variability of weekly influenza detections within the study period.

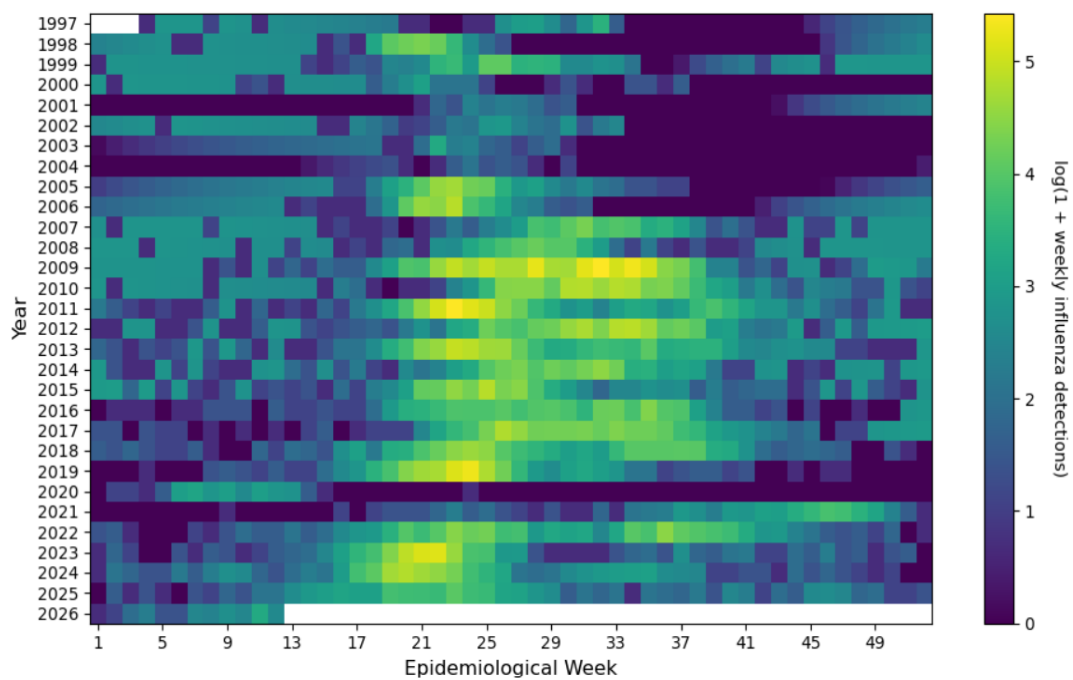


Figure 5. Annual–Seasonal Heatmap of Weekly Influenza Detections in South Africa, 1997–2026

The heatmap confirmed recurring seasonal concentrations while highlighting year-to-year differences in peak magnitude and timing. Reduced reported activity during the COVID-19 period was also visible, indicating that the evaluation series contains a structurally unusual period in addition to its usual seasonal cycle.

3.2. Stationarity and SARIMA Diagnostics

The ADF test produced a statistic of -8.78 with $p = 2.35 \times 10^{-14}$, leading to rejection of the unit-root null at the 0.05 level and supporting $d = 0$ for non-seasonal differencing. Pronounced dependence near lag 52 motivated a seasonal period of 52 weeks. Seasonal differencing was considered separately from the non-seasonal stationarity decision and

was retained in the final specification after candidate-model comparison. The ACF showed strong short-term autocorrelation together with recurring dependence around the 52-week lag, while the PACF showed a dominant short-term spike consistent with a low-order autoregressive component (Figures 6a and 6b). The final SARIMA model captured these temporal features while using seasonal differencing to represent the annual cycle.

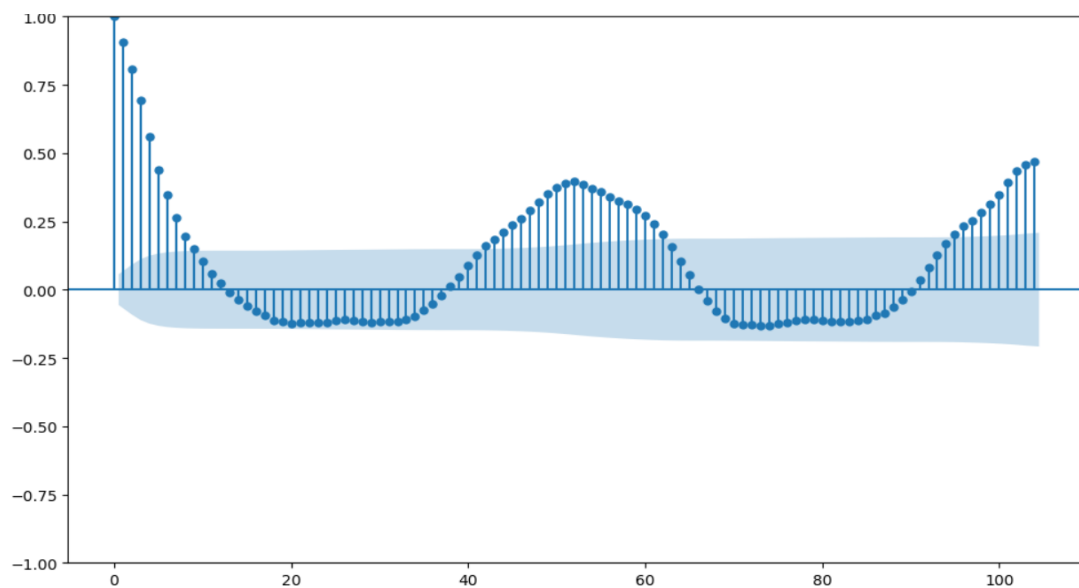


Figure 6a. ACF Plot of Weekly Influenza Detections: Training Series

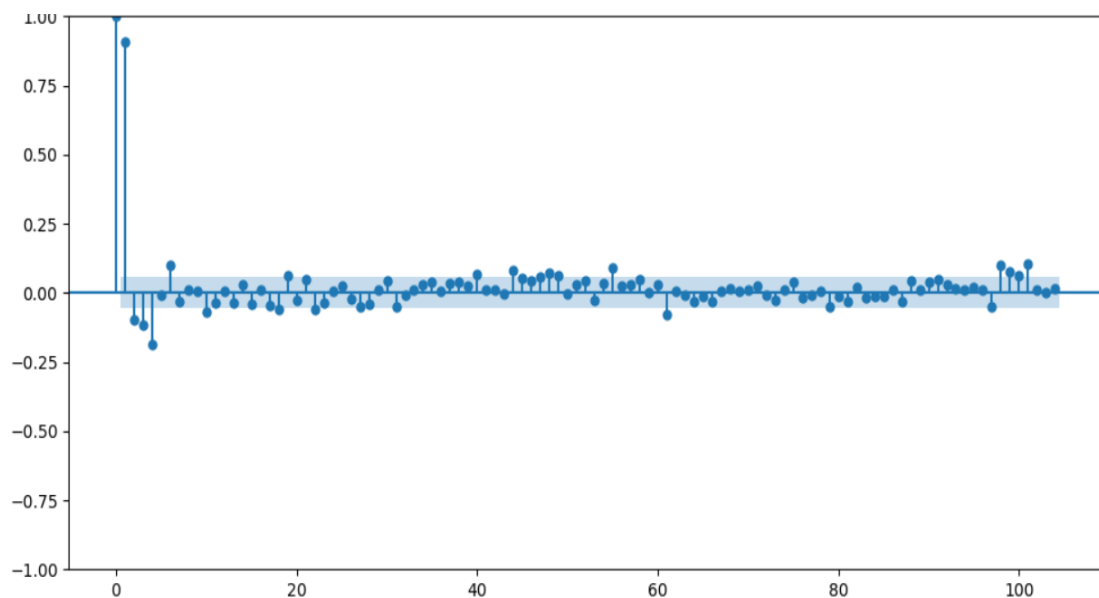


Figure 6b. PACF Plot of Weekly Influenza Detections: Training Series

The seasonally differenced plots in Figures 7a and 7b were used to assess the remaining dependence after removing the annual cycle and supported the low-order seasonal specification used in the final SARIMA model.

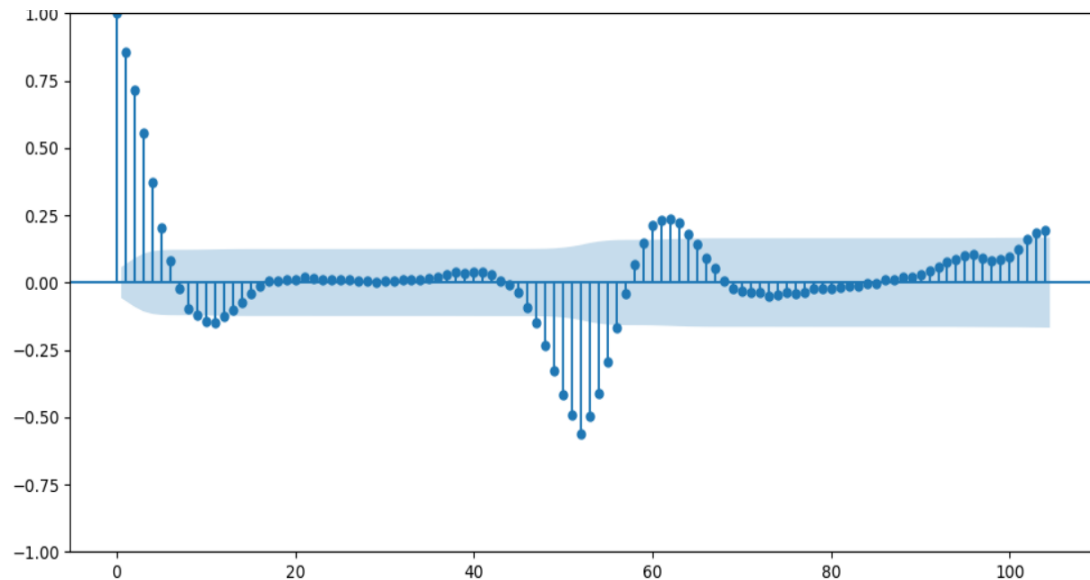


Figure 7a. ACF of the Seasonally Differenced Weekly Influenza Detection Series (52-week seasonal difference)

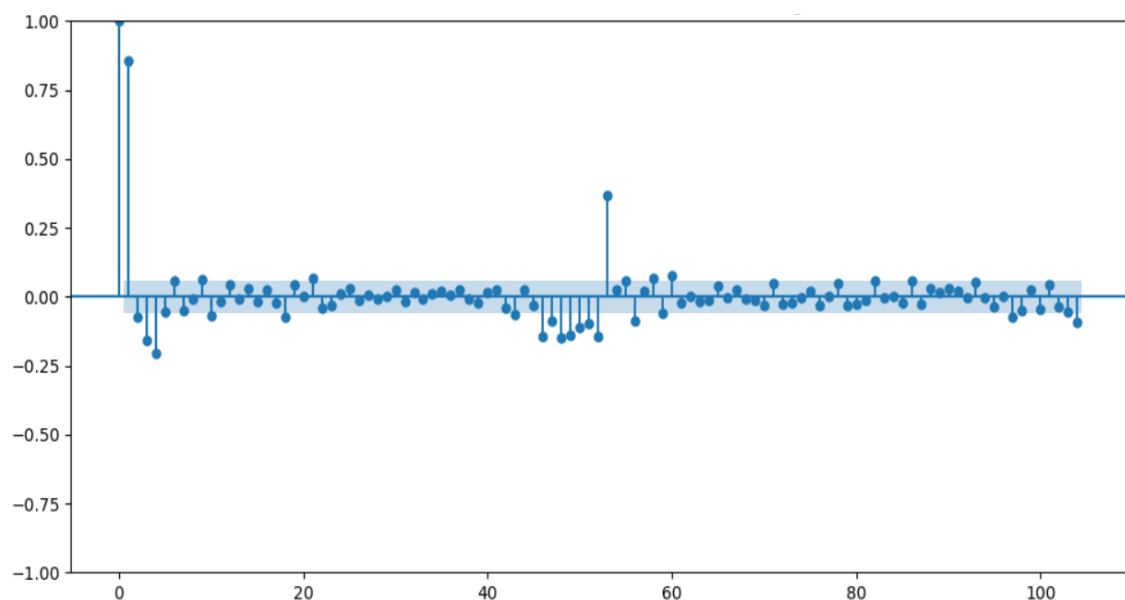


Figure 7b. PACF of the Seasonally Differenced Weekly Influenza Detection Series (52-week seasonal difference)

The selected SARIMA(1,0,1)×(1,1,1)₅₂ model estimated significant AR(1) ($\beta = 0.8633$, $p < 0.001$) and MA(1) ($\beta = 0.0480$, $p = 0.002$) terms. The seasonal AR term at lag 52 was not significant ($\beta = -0.0001$, $p = 0.996$), whereas the seasonal MA term was significant ($\beta = -0.8559$, $p < 0.001$). The model produced log-likelihood -4407.230, AIC 8824.459, BIC 8849.524, and HQIC 8833.937. The SARIMA model estimation output is shown in Table 3 below.

Table 3. SARIMA model estimation output

Parameter/Statistic	Estimate	Std. Error	z-value	p-value
AR(1)	0.8633	0.007	124.334	0.000
MA(1)	0.0480	0.015	3.117	0.002
Seasonal AR (52)	-0.0001	0.022	-0.006	0.996
Seasonal MA (52)	-0.8559	0.015	-56.735	0.000
Log-Likelihood	-4407.230			
AIC	8824.459			
BIC	8849.524			
HQIC	8833.937			

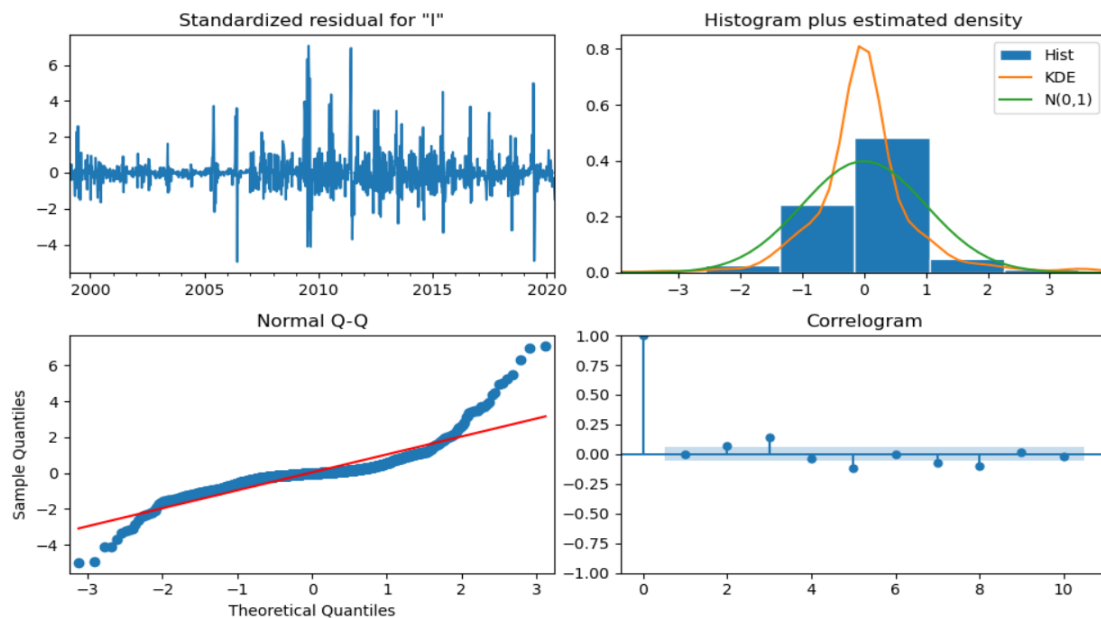


Figure 8. Diagnostic plots for standardized residuals of the fitted SARIMA model

The Residual diagnostics indicated departures from ideal residual behavior. The Ljung–Box test suggested remaining serial dependence at the tested lags, the Jarque–Bera test

indicated non-normality with heavy-tailed residuals, and the heteroskedasticity test indicated time-varying residual variance. The residual plots in Figure 8 showed that the series was centered approximately around zero but included larger deviations and spikes. These departures do not preclude point forecasting but should be considered when interpreting uncertainty estimates.

3.3. LSTM Training Results

The training and validation losses declined during model fitting, with early stopping used to limit unnecessary additional training once validation performance stopped improving. The final network used the 52-week input window and the two recurrent layers described in Section 2.4.2. The LSTM architecture in Table 4 comprised two stacked recurrent layers with 64 and 32 units, respectively, using the hyperbolic tangent (tanh) activation function.

Table 4. The LSTM Architecture Table

Layer (type)	Configuration	Output Shape	Param #
LSTM layer 1	64 units, tanh, return sequences	(None, 52, 64)	16,896
Dropout 1	20%	(None, 52, 64)	0
LSTM layer 2	32 units, tanh	(None, 32)	12,416
Dropout 2	20%	(None, 32)	0
Dense output layer	1 unit	(None, 1)	33
Total			29,345 (114.63KB)

The first LSTM layer returned the complete 52-step sequence, whereas the second returned the final 32-dimensional representation. Each LSTM layer was followed by dropout layers with a rate of 0.20 to reduce overfitting. A single-unit Dense output layer generated the one-step-ahead influenza detection forecast. The network used tanh activation in the LSTM layers and contained 29,345 (114.63KB) trainable parameters, with no non-trainable parameters (0B). The network was optimized using the Adam optimizer with a learning rate of 0.001 and mean squared error (MSE) as the loss function. The training used a batch size of 32 and a maximum of 100 epochs (Figure 9). Ten percent of the training sequences were reserved for validation, and early stopping with a patience of 10 epochs was applied based on validation loss, with the best model weights restored.

A random seed of 42 was specified to improve reproducibility, and no random shuffling was performed during training to preserve temporal ordering.

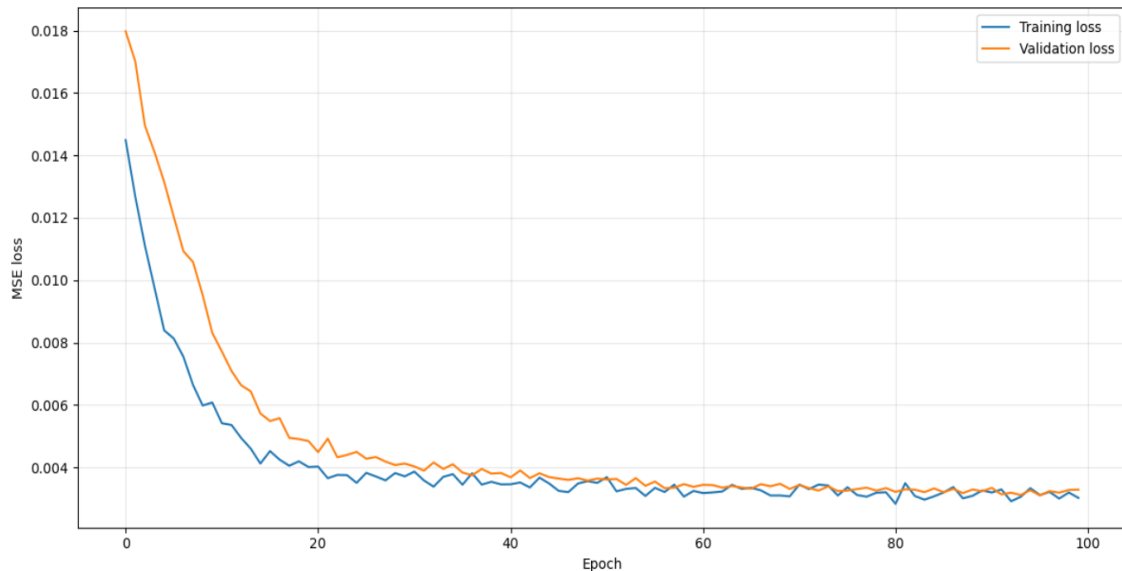


Figure 9. LSTM Training and Validation Loss

3.4. Forecasting Performance

The combined forecast plot in Figure 10 showed that all three approaches reproduced part of the recurring annual pattern, but their short-term trajectories differed. The Seasonal Naïve forecasts followed the corresponding observations from the previous seasonal cycle and therefore diverged when the magnitude or timing of activity changed. SARIMA captured the broad seasonal structure but produced smoother forecasts and underestimated several high-detection peaks. LSTM tracked the observed series more closely across much of the test period, including several major peaks and short-term fluctuations.

The Seasonal Naïve benchmark produced RMSE = 29.93 and MAE = 16.02. SARIMA improved on the benchmark with RMSE = 23.58 and MAE = 15.56. LSTM produced RMSE = 9.49 and MAE = 6.31. Relative to SARIMA, these values correspond to reductions of 59.8% in RMSE and 59.4% in MAE. Relative to Seasonal Naïve, the reductions were 68.3% and 60.6%, respectively. The results are summarized in Table 5.

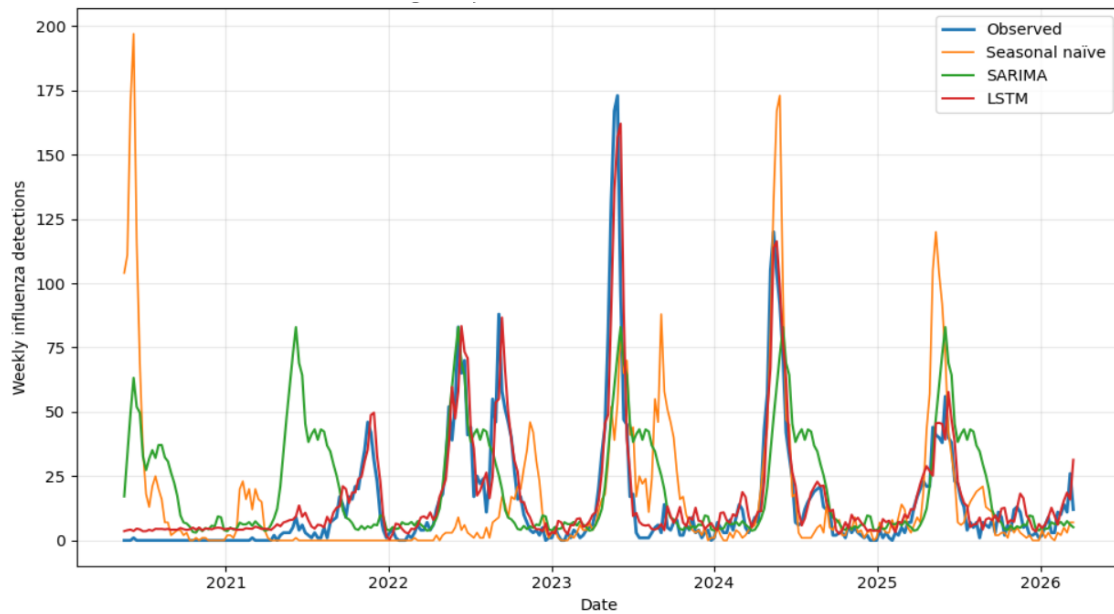


Figure 10. Forecasting performance of Seasonal Naïve, SARIMA, and LSTM models

Table 5. Performance evaluation of forecasting models

Metric	Seasonal Naïve	SARIMA	LSTM	Interpretation
RMSE	29.93	23.58	9.49	LSTM produced the lowest overall forecast error and was least affected by large errors.
MAE	16.02	15.56	6.31	LSTM had substantially lower average absolute forecast error.
Coverage_95(%)	94.43	94.75	95.74	All three of the developed models have achieved close to the desired 95% coverage; LSTM has reached slightly higher than the desired level, while Seasonal Naïve and SARIMA have reached slightly lower than the desired level.
Mean_Interval_Width	124.81	70.68	27.44	LSTM produced the narrowest predictive intervals while maintaining coverage close to the nominal level

The result of the Diebold–Mariano test for SARIMA and LSTM forecasts in Table 6 showed that the test yielded statistics of 6.6661 ($p < 0.001$) under MSE loss and 9.7569 ($p < 0.001$) under MAE loss; because the loss differential was SARIMA minus LSTM, the positive statistics indicate greater forecast loss for SARIMA over the evaluated test period.

Table 6: Diebold–Mariano test for SARIMA and LSTM forecasts

Comparison	Loss Function	DM statistic	p-value	Interpretation
SARIMA vs LSTM	MSE	6.6661	<0.001	Significant difference
SARIMA vs LSTM	MAE	9.7569	<0.001	Significant difference

3.5. Prediction-Interval Performance

Figure 11 below presents the forecasts generated by the Seasonal Naïve, SARIMA, and LSTM models, alongside their corresponding 95% prediction intervals. Empirical 95% interval coverage was 94.43% for Seasonal Naïve, 94.75% for SARIMA, and 95.74% for LSTM. Mean interval widths were 124.81, 70.68, and 27.44, respectively. Thus, all three approaches achieved coverage close to the nominal 95% level, while the LSTM intervals were substantially narrower. These interval results are descriptive for the evaluated holdout because the uncertainty procedures were model-specific rather than assessed through a common rolling-origin framework.

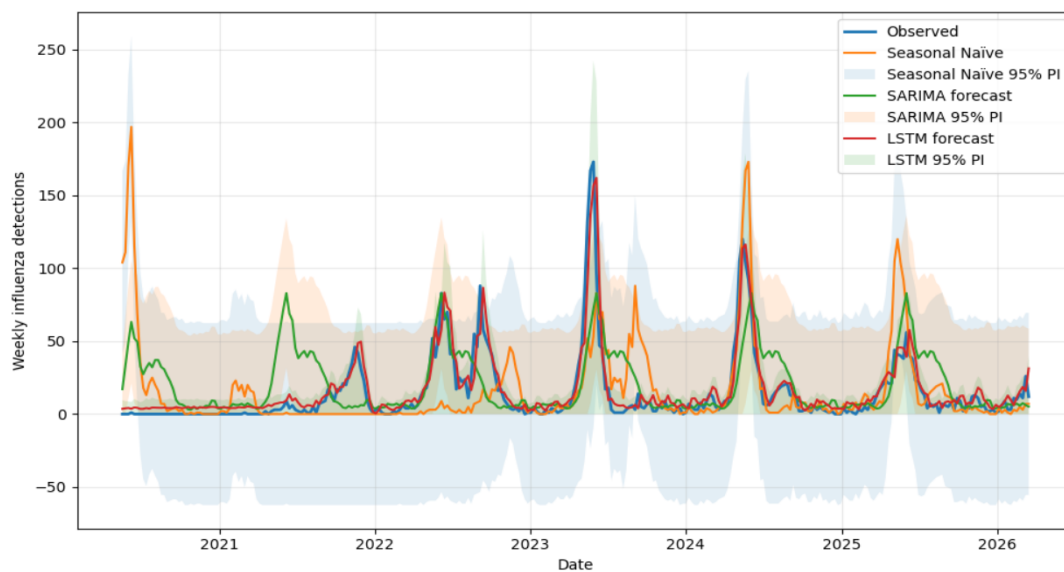


Figure 11. Comparison of Seasonal Naïve, SARIMA, and LSTM forecasts with 95% prediction intervals

3.6. Discussion

The observed performance differences among the forecasting approaches can be interpreted in light of their underlying methodological characteristics. SARIMA is a linear statistical time-series framework that models autoregressive and moving-average dependence together with seasonal structure after appropriate non-seasonal and seasonal differencing. This structure makes SARIMA particularly suitable for series exhibiting stable autocorrelation and recurring seasonal patterns. Previously studies have demonstrated the efficacy of SARIMA and related seasonal ARIMA models in the prediction of infectious disease, including influenza and tuberculosis, particularly where a periodic pattern is present [20], [21], [22]. However, linear seasonal models may exhibit reduced forecasting accuracy in the presence of abrupt series fluctuations, structural breaks, irregular fluctuations, and complex temporal dependencies [33].

LSTM, by contrast, is a recurrent deep-learning architecture designed to learn sequential dependencies from historical observations. Its ability to represent complex temporal relationships provides a potential advantage for epidemiological time series whose observed dynamics may vary substantially over time. Previous studies have reported cases in which LSTM and other deep-learning approaches outperform conventional statistical models for infectious-disease forecasting, including influenza and COVID-19 [28], [31], [35]. In the present study, LSTM achieved substantially lower RMSE and MAE than SARIMA, with an RMSE of 9.49 and MAE of 6.31, compared with 23.58 and 15.56, respectively, for SARIMA. These results indicate that LSTM provided more accurate forecasts of the evaluated weekly laboratory influenza detection series under the current experimental setup. However, the lower predictive errors of LSTM do not, by themselves, demonstrate that the underlying influenza detection process is nonlinear. No formal statistical test of nonlinearity, such as the BDS test, was conducted in the present study. Therefore, the superior LSTM performance should be interpreted as evidence of better predictive performance under the present modelling configuration rather than as definitive evidence of nonlinear epidemiological dynamics [25], [30]. Studies leveraging digital surveillance data and machine-learning approaches have similarly demonstrated the potential for improved predictive accuracy in influenza forecasting contexts [6], [38].

The inclusion of the Seasonal Naïve model provides an important benchmark for interpreting the performance of the more complex models. Seasonal Naïve forecasting

provides a simple persistence-based approach in which the value for a given week is forecast from the corresponding week in the previous seasonal cycle. In the present study, the Seasonal Naïve model produced an RMSE of 29.93 and MAE of 16.02, which were higher than the corresponding errors for both LSTM and SARIMA. The result indicates that simply reproducing the previous seasonal pattern was less effective than incorporating additional temporal information through SARIMA or LSTM. In particular, the lower RMSE and MAE of LSTM relative to Seasonal Naïve demonstrate that its performance advantage was not merely a consequence of comparing a complex machine-learning model with a conventional statistical model. Similarly, SARIMA also improved upon the Seasonal Naïve benchmark, particularly in terms of RMSE. The Seasonal Naïve results therefore provide evidence that modelling temporal dependence beyond simple seasonal persistence can improve forecasting accuracy for the South African laboratory influenza detection series.

The formal statistical comparison of SARIMA and LSTM models based on the results of the previous step also highlights the differences in their accuracy in terms of MSE and MAE. Particularly, when MSE served as a loss function, the DM-test statistic equals 6.6661 (p -value < 0.001). By contrast, when MAE was used, the test statistic was 9.7569 (p -value < 0.001). Thus, both values are statistically significant at the 1% level, implying that the forecast errors of the two models are different. Given the sign of the test statistic, it is evident which model has lower forecast errors. Therefore, the initial differences in terms of RMSE and MAE are confirmed statistically and not due to chance. However, it is important to note that these results are based on the particular holdout period used to evaluate the models. While the DM-test indicates a statistically significant difference in accuracy between SARIMA and LSTM for the selected forecast horizon and loss functions, this conclusion applies only to the specific holdout sample, it does not claim that LSTM is universally superior to SARIMA across different datasets, forecasting horizons, or surveillance environments.

The results also have important implications for the interpretation of the forecast trajectories. The LSTM forecasts reproduced more of the observed short-term fluctuations and high-detection peaks, while the SARIMA forecasts were smoother and underestimated the magnitude of some high-detection periods. By construction, the Seasonal Naïve forecasts reproduced the seasonal pattern of the previous cycle but were

less reactive to deviations from this historical pattern. The differences match those of the numerical error results, as LSTM achieved the lowest RMSE and MAE of the three models evaluated. The ability of LSTM to follow short-term variation more closely may be especially useful in surveillance applications to anticipate changes in laboratory detection. However, forecast accuracy should not be interpreted directly as outbreak-detection capability, since the study did not evaluate operational alert thresholds or outbreak-warning performance.

While LSTM has better accuracy than SARIMA, the latter remains an important and useful forecasting approach. Its transparent parameterization has an interpretable representation of temporal and seasonal dependence and requires relatively low computational resources. The Seasonal Naïve benchmark also highlights that SARIMA provides additional value in forecasting beyond the simple seasonal persistence. Thus, while LSTM achieved the best point forecasts in the present evaluation, SARIMA remains valuable as an interpretable statistical benchmark and may be preferable in contexts where simplicity, transparency, or computational efficiency are priorities. The findings, therefore, suggest a context-dependent perspective on model selection rather than suggesting predictive accuracy as the sole criterion for choosing a forecasting approach.

The predictive-uncertainty results provide an additional perspective on the performance of SARIMA and LSTM. Both models achieved empirical coverage close the nominal 95%, with LSTM achieving 95.74% coverage compared to 94.75% for SARIMA. LSTM forecasts also produced substantially narrower predictive intervals under the evaluation procedure, with an average width of 27.44 compared to 70.68 for SARIMA. These results suggest that the LSTM forecasts achieved close to nominal coverage with narrower intervals under the evaluation procedure used. However, narrower prediction intervals should not automatically be assumed to be better uncertainty estimates unless sufficient coverage is maintained. Furthermore, the uncertainty estimates were obtained using procedures specific to each model and were not evaluated within a shared rolling-origin framework. Therefore, the interval findings should be regarded as model-specific evidence of forecast uncertainty over the evaluated holdout period rather than conclusive proof that LSTM provides better-calibrated uncertainty forecasts in general.

From a public-health surveillance perspective, the results suggest that machine-learning methods such as LSTM could provide valuable forecasting support for weekly laboratory influenza detection surveillance. More accurate forecasts might help identify periods of heightened detection and guide surveillance planning, resource allocation, and preparedness. However, the phrase “laboratory influenza detections” is critically important in interpreting the results. WHO FluNet data represent laboratory-confirmed detections reported via surveillance systems and should not be treated as direct measures of influenza incidence at the population level. Shifts in testing intensity, specimen availability, surveillance coverage, reporting completeness, healthcare-seeking behavior, and other surveillance processes can affect the detection counts observed. Consequently, the forecasts generated in this study should be understood as forecasts of reported laboratory influenza detections rather than forecasts of the true underlying incidence of influenza in the population.

An important consideration in interpreting the forecasting results is the potential effect of the COVID-19 pandemic on influenza surveillance and the underlying detection process. The fixed holdout period begins in May 2020 and therefore includes observations from the pandemic and post-pandemic periods, during which influenza circulation, healthcare-seeking behavior, healthcare utilization, testing practices, and surveillance activities may have differed from pre-pandemic conditions. These changes could have altered the temporal distribution and reporting of laboratory influenza detections and may consequently have affected model performance. The results should therefore be interpreted as performance on the observed surveillance series during the specified evaluation period rather than as evidence of generalizable predictive performance under unchanged surveillance conditions [58]. Future research should examine model performance separately across pre-pandemic, pandemic, and post-pandemic periods and consider surveillance-related covariates or regime-change approaches where appropriate.

Despite having taken the Seasonal Naïve benchmark into account, there are still some limitations to this study which should be noted. First, the analysis on the FluNet laboratory detections reports from South Africa, without incorporating population denominators or other indicators of surveillance intensity. Therefore, the differences in testing volume, availability of specimen, level of surveillance, and reporting practices

could account for the observed differences in the series. Second, the models relied exclusively on historical detection counts and did not incorporate potentially informative exogenous predictors. Third, although the fixed chronological holdout provides a genuine out-of-sample evaluation using the same observations for all models, the study did not use rolling-origin or expanding-window validation across multiple forecast origins. A single holdout period may not fully capture how model performance changes across different epidemiological regimes. In addition, the present analysis did not comprehensively evaluate multiple forecasting horizons. Finally, although prediction intervals were assessed for SARIMA and LSTM, their calibration and sharpness were not evaluated under a common rolling-origin framework.

Future research should therefore evaluate the robustness of these findings using expanding-window rolling-origin validation and multiple forecasting horizons. Additional established forecasting approaches, including exponential smoothing and other epidemiological forecasting benchmarks, could also be incorporated to determine whether the additional complexity of LSTM provides consistent incremental predictive value. Formal assessment of nonlinearity would further help clarify whether the improved LSTM performance is associated with nonlinear temporal structure in the detection series. Incorporating exogenous information such as weather conditions, Google Trends, vaccination coverage, school calendars, population mobility, specimen-processing volume, and influenza positivity rates could provide additional explanatory and predictive information while helping distinguish changes in surveillance intensity from changes in influenza activity. Future studies could also investigate interpretable hybrid models that combine statistical and deep-learning approaches and evaluate forecasting systems using operational warning criteria to determine their practical value for public-health surveillance and early-warning applications.

The findings indicate that LSTM provided the best point-forecasting performance among the three approaches evaluated, achieving lower RMSE and MAE than both SARIMA and the Seasonal Naïve benchmark. The statistically significant DM results further demonstrated a significant difference in forecasting accuracy between LSTM and SARIMA under both MSE and MAE loss functions. These findings support the potential value of LSTM for forecasting weekly laboratory influenza detections in South Africa under the present fixed-holdout evaluation design. Nonetheless, the findings should be interpreted

as relative performance indications based on the particular dataset, modelling setup and evaluation timeframe adopted in this study, rather than proof of overall superiority. Additional research is needed, encompassing broader rolling-origin validation, further benchmarks, multiple forecasting horizons, a rigorous assessment of nonlinearity, and relevant surveillance and epidemiological predictors, to verify these results.

4. CONCLUSION

This study compared how well the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) models forecasted weekly influenza detection counts in South Africa, using WHO FluNet surveillance data with a Seasonal Naïve model also included as a benchmark. The findings show that SARIMA captured the recurring seasonal pattern in the surveillance series, while the Seasonal Naïve model acted as a simple seasonal persistence baseline against which the predictive gains of the more complex models could be judged. During the fixed chronological test period, the LSTM yielded lower forecasting errors than the chosen SARIMA and Seasonal Naïve specifications, and the Diebold–Mariano test offered statistical evidence that forecast accuracy differed between SARIMA and LSTM under the forecasting design that was evaluated. These findings, however, pertain only to the single holdout period and should not be taken as proof that one model is generally superior. The analysis relied on surveillance-based influenza detection counts instead of population-level incidence and made no adjustment for population or surveillance denominators, such as the number of specimens tested; consequently, the observed changes may also mirror shifts in testing intensity, specimen collection, reporting practices, surveillance coverage, healthcare-seeking behavior, and other aspects of the detection process. No exogenous predictors were incorporated, and the model-specific uncertainty estimates were not assessed within a shared rolling-origin framework. Thus, future research should employ rolling-origin and multi-horizon validation, compare these models with other well-established forecasting approaches, include relevant exogenous and surveillance-related variables such as weather, vaccination coverage, school calendars, mobility, specimen-processing volumes, and positivity rates, and evaluate population- or surveillance-adjusted measures when suitable denominator data exist. Subsequent work should also explore interpretable hybrid methods and assess forecasting models against operational outbreak-warning

criteria to establish their practical value for influenza surveillance and early-warning applications.

REFERENCES

- [1] World Health Organization, "Influenza (seasonal)," Feb. 28, 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/influenza-%28seasonal%29>.
- [2] F. E. L. Guma, "Analysis of influenza-like illness trends in Saudi Arabia: A comparative study of statistical and deep learning techniques," *Osong Public Health Res. Perspect.*, vol. 16, no. 3, pp. 270–284, 2025, doi: 10.24171/j.phrp.2025.0080.
- [3] S. Fiandrino, A. Bizzotto, G. Guzzetta, S. Merler, F. Baldo, E. Valdano, A.M. Urdiales, A. Bella, F. Celino, L. Zino, and A. Rizzo, "Collaborative forecasting of influenza-like illness in Italy: The Influcast experience," *Epidemics*, vol. 50, Art. no. 100819, 2025, doi: 10.1016/j.epidem.2025.100819.
- [4] E. O. Nsoesie, O. Oladeji, A. S. A. A. Abah, and M. L. Ndeffo-Mbah, "Forecasting influenza-like illness trends in Cameroon using Google Search Data," *Sci. Rep.*, vol. 11, Art. no. 6713, 2021, doi: 10.1038/s41598-021-85987-9.
- [5] National Institute for Communicable Diseases, "Influenza surveillance report 2025," 2026, [Online]. Available: <https://www.nicd.ac.za/influenza-surveillance-report-2025>
- [6] S. O. Olukanmi, F. V. Nelwamondo, and N. I. Nwulu, "Leveraging Google search data and artificial intelligence methods for provincial-level influenza forecasting: A South African case study," *Int. J. Online Biomed. Eng.* 18, no. 11 2022, doi: 10.3991/ijoe.
- [7] S. B. Choi and I. Ahn, "Forecasting seasonal influenza-like illness in South Korea after 2 and 30 weeks using Google Trends and influenza data from Argentina," *PLoS ONE*, vol. 15, no. 7, Art. no. e0233855, 2020, doi: 10.1371/journal.pone.0233855.
- [8] S. Cho, C. H. Sohn, M. W. Jo, S.-Y. Shin, J. H. Lee, S. M. Ryoo, W. Y. Kim, and D.-W. Seo, "Correlation between national influenza surveillance data and Google Trends in South Korea," *PLoS ONE*, vol. 8, no. 12, Art. no. e81422, 2013, doi: 10.1371/journal.pone.0081422.
- [9] M. Kang, H. Zhong, J. He, S. Rutherford, and F. Yang, "Using Google Trends for influenza surveillance in South China," *PLoS ONE*, vol. 8, no. 1, Art. no. e55205, 2013, doi: 10.1371/journal.pone.0055205.

- [10] C. W. Chukwu, G. Obaido, I. D. Mienye, K. Aruleba, E. Esenogho, and C. Modisane, "Time series modeling of monkeypox incidence in Central Africa's endemic regions," *Mach. Learn. Appl.*, vol. 22, Art. no. 100778, 2025, doi: 10.1016/j.mlwa.2025.100778.
- [11] Y. Teng, D. Bi, G. Xie, Y. Jin, Y. Huang, B. Lin, X. An, D. Feng, and Y. Tong, "Dynamic forecasting of Zika epidemics using Google Trends," *PLoS ONE*, vol. 12, no. 1, Art. no. e0165085, 2017, doi: 10.1371/journal.pone.0165085.
- [12] E. G. Igwe and W. M. Olatokun, "Patterns and conversational structure of knowledge sharing on Ebola virus disease among healthcare practitioners: Observations from a pilot study," *J. Librarianship Inf. Sci.*, vol. 55, no. 3, pp. 836–847, 2023, doi: 10.1177/09610006221106180.
- [13] C. Alicino, N. L. Bragazzi, V. Faccio, D. Amicizia, D. Panatto, R. Gasparini, G. Icardi, and A. Orsi, "Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes," *Infect. Dis. Poverty*, vol. 4, Art. no. 54, 2015, doi: 10.1186/s40249-015-0090-9
- [14] S. A. Ebiaredoh-Mienye, T. G. Swart, E. Esenogho, and I. D. Mienye, "A machine learning method with filter-based feature selection for improved prediction of chronic kidney disease," 2022, doi: 10.3390/bioengineering9080350.
- [15] M. Effenberger, A. Kronbichler, J. I. Shin, G. Mayer, H. Tilg, and P. Perco, "Association of the COVID-19 pandemic with Internet search volumes: A Google Trends analysis," *Int. J. Infect. Dis.*, vol. 95, pp. 192–197, 2020, doi: 10.1016/j.ijid.2020.04.033.
- [16] S. M. Ayyoubzadeh, S. M. Ayyoubzadeh, H. Zahedi, M. Ahmadi, and S. R. Niakan Kalhori, "Predicting COVID-19 incidence through analysis of Google Trends data in Iran: Data mining and deep learning pilot study," *JMIR Public Health Surveill.*, vol. 6, no. 2, Art. no. e18828, 2020, doi: 10.2196/18828.
- [17] C. Li, L. J. Chen, X. Chen, M. Zhang, C. P. Pang, and H. Chen, "Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020," *Euro Surveill.*, vol. 25, no. 10, Art. no. 2000199, 2020, doi: 10.2807/1560-7917.ES.2020.25.10.2000199.
- [18] X. Zhou, J. Ye, and Y. Feng, "Tuberculosis surveillance by analyzing Google Trends," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2247–2254, 2011, doi: 10.1109/TBME.2011.2132132.
- [19] A. Mavragani and G. Ochoa, "Infoveillance of infectious diseases in USA: STDs, tuberculosis, and hepatitis," *J. Big Data*, vol. 5, Art. no. 30, 2018, doi: 10.1186/s40537-018-0140-9.

- [20] S. Chadsuthi, S. Iamsirithaworn, W. Triampo, and C. Modchang, "Modeling seasonal influenza transmission and its association with climate factors in Thailand using time-series and ARIMAX analyses," *Comput. Math. Methods Med.*, vol. 2015, Art. no. 436495, 2015, doi: 10.1155/2015/436495.
- [21] Q. Mao, K. Zhang, W. Yan, and C. Cheng, "Forecasting the incidence of tuberculosis in China using the seasonal auto-regressive integrated moving average (SARIMA) model," *J. Infect. Public Health*, vol. 11, no. 5, pp. 707–712, 2018, doi: 10.1016/j.jiph.2018.04.009.
- [22] T. Xiao-Qing, Z. Zhan-Lin, G. Zheng, Y. Mahan, H. Bing-Xue, T. Tian, A. Ainiwaer, C. Zhen, G. Hailili, F. Xu-Cheng, and D. Jiang-Hong, "Forecasting influenza-like illness in Urumqi based on ARIMAX model," *Chin. J. Dis. Control Prev.*, vol. 22, no. 6, pp. 590–593, 2018, doi: 10.16462/J.CNKI.ZHJBKZ.2018.06.012.
- [23] W. Anggraeni and L. Aristiani, "Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method: Case study Surabaya, Indonesia," in *Proc. Int. Conf. Inf. Commun. Technol. Syst. (ICTS)*, 2016, doi: 10.1109/ICTS.2016.7910283
- [24] C. Poirier, A. Lavenue, V. Bertaud, B. Campillo-Gimenez, E. Chazard, M. Cuggia, and G. Bouzillé, "Real time influenza monitoring using hospital big data in combination with machine learning methods: Comparison study," *JMIR Public Health Surveill.*, vol. 4, no. 4, Art. no. e11361, 2018, doi: 10.2196/11361.
- [25] M. Santillana, A. T. Nguyen, M. Dredze, M. J. Paul, E. O. Nsoesie, and J. S. Brownstein, "Combining search, social media, and traditional data sources to improve influenza surveillance," *PLoS Comput. Biol.*, vol. 11, no. 10, Art. no. e1004513, 2015, doi: 10.1371/journal.pcbi.1004513.
- [26] S. Prasanth, U. Singh, A. Kumar, V. A. Tikkiwal, and P. H. J. Chong, "Forecasting spread of COVID-19 using Google Trends: A hybrid GWO-deep learning approach," *Chaos Solitons Fractals*, vol. 142, Art. no. 110336, 2021, doi: 10.1016/j.chaos.2020.110336.
- [27] S. Volkova, E. Ayton, K. Porterfield, and C. D. Corley, "Forecasting influenza-like illness dynamics for military populations using neural networks and social media," *PLoS ONE*, vol. 12, no. 12, Art. no. e0188941, 2017, doi: 10.1371/journal.pone.0188941.
- [28] C. T. Yang Y.-A. Chen, Y.-W. Chan, C.-L. Lee, Y.-T. Tsan, W.-C. Chan, and P.-Y. Liu, "Influenza-like illness prediction using a long short-term memory deep learning model with multiple open data sources," *J. Supercomput.*, vol. 76, pp. 9303–9329, 2020, doi: 10.1007/s11227-020-03182-5.

- [29] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data," *Int. J. Environ. Res. Public Health*, vol. 15, no. 8, Art. no. 1596, 2018, doi: 10.3390/ijerph15081596.
- [30] Y. T. Tsan, D. Y. Chen, P. Y. Liu, E. Kristiani, K. L. P. Nguyen, and C. T. Yang, "The prediction of influenza-like illness and respiratory disease using LSTM and ARIMA," 2022, doi: 10.3390/ijerph19031858.
- [31] D. Zhao, H. Zhang, Q. Cao, Z. Wang, S. He, M. Zhou, and R. Zhang, "The research of ARIMA, GM(1,1), and LSTM models for prediction of TB cases in China," *PLoS ONE*, vol. 17, no. 2, Art. no. e0262734, 2022, doi: 10.1371/journal.pone.0262734.
- [32] J. Shaman and A. Karspeck, "Forecasting seasonal outbreaks of influenza," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 50, pp. 20425–20430, 2012, doi: 10.1073/pnas.1208772109.
- [33] M. Riaz, M. H. Sial, S. Sharif, and Q. Mehmood, "Epidemiological forecasting models using ARIMA, SARIMA, and Holt–Winter multiplicative approach for Pakistan," *Comput. Math. Methods Med*, vol. 2023, Art. no. 8907610, 2023, doi: 10.1155/2023/8907610.
- [34] A. Watad, S. Watad, N. Mahroum, K. Sharif, H. Amital, N.L. Bragazzi, and M. Adawi, "Forecasting the West Nile Virus in the United States: An extensive novel data streams–based time series analysis and structural equation modeling of related digital searching behavior," *JMIR Public Health Surveill*, 2019, doi: 10.2196/publichealth.
- [35] V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos Solitons Fractals*, vol. 135, Art. no. 109864, 2020, doi: 10.1016/j.chaos.2020.109864.
- [36] F. E. L. Guma, "Comparative analysis of time series prediction models for visceral leishmaniasis: Based on SARIMA and LSTM," *Appl. Math. Inf. Sci*, vol. 18, no. 1, pp. 125–132, 2024, doi: 10.18576/amis/180113.
- [37] E. Budgell, A. L. Cohen, J. M. McAnerney, S. Walaza, S. A. Madhi, L. H. Blumberg, H. Dawood, K. Kahn, S. Tempia, and M. Venter, "Evaluation of two influenza surveillance systems in South Africa," *PLoS ONE*, 2015.
- [38] S. O. Olukanmi, F. V. Nelwamondo, and N. I. Nwulu, "Utilizing Google Search Data with deep learning, machine learning and time series modeling to forecast influenza-like illnesses in South Africa," *IEEE Access*, vol. 9, pp. 126822–126836, 2021, doi: 10.1109/ACCESS.2021.3110972.

- [39] S. O. O. and F. V. Nelwamondo, "Digital influenza surveillance: The prospects of Google Trends data for South Africa," 2020, doi: 10.1109/icABCD49160.
- [40] World Health Organization, "Global Influenza Surveillance and Response System (GISRS)," 2026. [Online]. Available: <https://www.who.int/initiatives/global-influenza-surveillance-and-response-system>.
- [41] World Health Organization, "FluNet," 2026. [Online]. Available: <https://www.who.int/tools/flunet>
- [42] A. B. Suthar, J. Wang, V. Seffren, R. E. Wiegand, S. Griffing, and E. Zell, "Public health impact of COVID-19 vaccines in the US: Observational study," *BMJ*, vol. 377, Art. no. e069317, 2022, doi: 10.1136/bmj-2021-069317.
- [43] A. K. Dubey, A. Kumar, V. García-Díaz, A. K. Sharma, and K. Kanhaiya, "Study and analysis of SARIMA and LSTM in forecasting time series data," *Sustain. Energy Technol. Assess.*, vol. 47, Art. no. 101474, 2021, doi: 10.1016/j.seta.2021.101474.
- [44] S. Aryal, D. Nadarajah, P. L. Rupasinghe, C. Jayawardena, and D. Kasthurirathna, "Comparative analysis of deep learning models for multi-step prediction of financial time series," *J. Comput. Sci.*, vol. 16, no. 10, pp. 1401–1416, 2020, doi: 10.3844/jcssp.2020.1401.1416.
- [45] Z. Zhao, M. Zhai, G. Li, X. Gao, W. Song, X. Wang, H. Ren, Y. Cui, Y. Qiao, J. Ren, L. Chen, and L. Qiu, "Study on the prediction effect of a combined model of SARIMA and LSTM based on SSA for influenza in Shanxi Province, China," *BMC Infect. Dis.*, vol. 23, Art. no. 71, 2023, doi: 10.1186/s12879-023-08025-1.
- [46] J. S. Adeyeye, and E. B. Nkemnole, "Predicting malaria incident using hybrid sarima-lstm model", *International Journal of Mathematical Sciences and Optimization: Theory and Applications*, 9(1), pp.59-80, 2023
- [47] P. Aghelpour, B. Mohammadi, and S. M. Biazar, "Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA," 2019: P. Aghelpour et al. *Theoretical and applied climatology*, 138(3), pp.1471-1480.
- [48] R. Zhao, J. Liu, Z. Zhao, M. Zhai, H. Ren, X. Wang, Y. Li, Y. Cui, Y. Qiao, J. Ren, L. Chen, and L. Qiu, "A hybrid model for tuberculosis forecasting based on empirical mode decomposition in China," *BMC Infect. Dis.*, vol. 23, Art. no. 665, 2023, doi: 10.1186/s12879-023-08609-x.
- [49] T. C. Mills, *A Very British Affair: Six Britons and the Development of Time Series Analysis During the 20th Century*. London, U.K.: Palgrave Macmillan, 2012.

- [50] K. E. ArunKumar, D. V. Kalaga, C. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, "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," *Alexandria Eng. J.*, vol. 61, no. 10, pp. 7585–7603, 2022, doi: 10.1016/j.aej.2022.01.011.
- [51] J. Yang, L. Yang, G. Li, J. Du, L. Ma, T. Zhang, X. Zhang, J. Yang, L. Feng, W. Yang, and C. Wang, "Comparative study on influenza time series prediction models in a megacity from 2010 to 2019 based on SARIMA and deep learning hybrid prediction model," 2022, doi: 10.21203/rs.
- [52] A. K. L. Bezerra, and É.M.C. Santos, "Prediction of daily confirmed COVID-19 cases in Sudan with ARIMA and Holt-Winters exponential smoothing," *Int. J. Dev. Res.*, vol. 10, no. 8, pp. 39408–39413, 2020, doi: 10.37118/ijdr.19811.08.2020.
- [53] L. Mamudu, A. Yahaya, and S. Dan'azumi, "Application of seasonal autoregressive integrated moving average (SARIMA) For Flows of River Kaduna," *Nigerian J. Eng.*, vol. 28, no. 2, pp. 84–90, 2021.
- [54] M. Othman, R. Indawati, A. A. Suleiman, M. B. Qomaruddin, and R. Sokkalingam, "Model forecasting development for dengue fever incidence in Surabaya City using time series analysis," *Processes*, vol. 10, no. 11, Art. no. 2454, 2022, doi: 10.3390/pr10112454.
- [55] A. H. Amshi, R. Prasad, and B. K. Sharma, "Forecasting cholera disease using SARIMA and LSTM models with discrete wavelet transform as feature selection," *J. Intell. Fuzzy Syst.*, vol. 45, no. 3, 2023, doi: 10.3233/JIFS-223901.
- [56] A. H. Amshi and R. Prasad, "Time series analysis and forecasting of cholera disease using discrete wavelet transform and seasonal autoregressive integrated moving average model," *Sci. Afr.*, vol. 20, Art. no. e01652, 2023, doi: 10.1016/j.sciaf. 2023.e01652.
- [57] H. Liu, C. Li, Y. Shao, X. Zhang, Z. Zhai, X. Wang, X. Qi, J. Wang, Y. Hao, Q. Wu, and M. Jiao, "Forecast of the trend in incidence of acute hemorrhagic conjunctivitis in China from 2011–2019 using the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Exponential Smoothing (ETS) models," *J. Infect. Public Health*, vol. 13, no. 2, pp. 287–294, 2020, doi: 10.1016/j.jiph.2019.12.008.
- [58] World Health Organization, "Influenza update No. 401," 2026, [Online]. Available: <https://www.who.int/publications/m/item/influenza-update-n-401>