# Machine Learning-Based Prediction of Seasonal Influenza Trends in Saudi Arabia: A Tool for Regional Public Health Planning

Alshaikh A. Shokeralla[1,*], Fathelrhman El Guma[1], Ali H. Abdalla[1], Amal E.Y. Hagsddig[2], Rahma AbuBakr Musa[3,4], M.A.M. Eltaweel[1,5], Abdelaziz H. Elawad[1] and Ibrahim Elshamy[6,7]

[1]*Department of Mathematics, Faculty of Science, Al-Baha University, Al-Aqiq 65931, Saudi Arabia*

[2]*Department of Statistics, Faculty of Engineering, Mashreq University, Khartoum, Sudan*

[3]*Biology Department, Faculty of Science, Al-Baha University, Saudi Arabia*

[4]*Biology Department, Faculty of Applied and Industrial Science, Bahri University, Sudan*

[5]*Faculty of Science Ain Shams University Cairo Egypt*

[6]*Higher Institute of Engineering & Technology, Almanzala, Egypt*

[7]*Mathematics Department, Faculty of Science, Al Baha University, KSA*

**Abstract:** Influenza has continued to be a worldwide social health problem, especially in high-density population areas with minimal early-warning mechanisms. The present study evaluates the predictive ability of two machine learning models Support Vector Regression (SVR) and Random Forest (RF) to predict weekly influenza cases in Saudi Arabia, spanning from 2017 to 2022, on 313 weekly influenza records in the WHO Global Influenza Surveillance and Response System (GISRS). The performance of these models was measured with $R^2$, MAE, MSE, and RMSE. Although SVR had a better training accuracy ($R^2$ = 0.96), RF had a better generalization ($R^2$ = 0.818) and more consistent predictions at the peaks of the seasons. These observations show that RF is appropriate to real-time influenza surveillance and can provide a replicable and versatile framework to assist data-driven epidemic preparedness in Saudi Arabia and other similar contexts throughout MENA and Asia-Pacific.

**Keywords:** Forecasting, Influenza, Machine learning, Random Forest, Saudi Arabia, Support Vector Regression.

## 1. INTRODUCTION

Seasonal influenza in the world is one of the leading causes of morbidity and mortality with an estimated 290,000 to 650,000 respiratory deaths every year [1]. It is transmitted by airborne droplets and is particularly common in highly populated zones, as well as among the vulnerable populations, such as children, older people and those with chronic conditions. Influenza surveillance is a social health concern in the Middle East where the variability of climatic conditions, cross-border travel, and urban congestion are making it challenging to contain the disease. In the Kingdom of Saudi Arabia (KSA), the frequent large-scale crowds like Hajj and Umrah increase the transmission risks, and well-developed forecasting tools are essential in early detection and interventions [2,3].

Early detection of influenza pandemics supports immunization measures, distribution of resources and preparedness to epidemics. Even though classical time-series forecasting models, including SARIMA, ARIMA, and Holt-Winters exponential smoothing, have been used to predict influenza, they generally assume stationarity and linearity. These assumptions make them less effective in the representation of the multivariate, nonlinear and complex nature of the transmission of infectious diseases [4,6]. Machine learning (ML) methods have become more versatile and data-driven alternatives to traditional public health datasets as they are growing more multidimensional and dynamic. Certain ML models, including Support Vector Regression (SVR) and Random Forest (RF), are capable of processing nonlinear patterns of data on a large scale and a multiplicity of predictors with increased robustness than classical models [7,9]. Recent reports have shown that ML can be used to predict influenza-like illness (ILI) cases in the countries of China and South Korea [10, 11]. However, these models have hardly been tested in resource changing, climate diversified or socio-culturally different environments like Saudi Arabia.

Although there is an increased interest in machine learning in forecasting of infectious disease outbreaks, its usage is hardly present in the Middle East countries, especially when it involves real-time multi-year national surveillance data. The comparative and region-specific evidence on model performance between the influenza seasons are lacking, and it is difficult to implement AI

*Address correspondence to this author at the Department of Mathematics, Faculty of Science, Al-Baha University, Al-Aqiq 65931, Saudi Arabia;
E-mail: sshokeralla@gmail.com

tools with confidence in public health systems. This study was carried out to validate and generalize ML framework applicable to Saudi Arabia, a region that faces special problems like mass gatherings and disease variability due to climate conditions. Moreover, there is little literature comparing the various machine learning models to long-term national epidemiological data, specifically with the goal to determine, among other things, not only the accuracy but also the ability to generalize across influenza seasons. Real-time forecasting requires models that would be effective in the dynamic viral strains, weather, population dynamics and delays in surveillance [12,13]. Influenza information in Saudi Arabia is reported to the WHO within the global surveillance systems; however, this information is hardly utilized in predictive modeling. This study was carried out to compare two machine learning models-SVR and RF to forecast weekly influenza cases using six years of surveillance data at the national level of surveillance (2017-2022). This research contributes to the advancement of regionally adaptable, ML-driven early warning systems and supports evidence-based decision-making in epidemic forecasting for Saudi Arabia and other Middle Eastern and Asia-Pacific countries seeking to modernize public health surveillance.

Therefore, the objectives of this research are outlined as follows: (1) to create and compare Support Vector Regression (SVR) and Random Forest (RF) models for weekly influenza case forecasting in Saudi Arabia using six years of national surveillance data (2017–2022). (2) to evaluate the prediction accuracy and generalizability of models across influenza seasons. And (3) to provide clues which can be used in combination with regional public health planning and early warning systems to improve future machine learning based forecast tools.

## 2. METHODS

### 2.1. Data Source and Description

The data were obtained from the World Health Organization's Global Influenza Surveillance and Response System (WHO GISRS), which provides publicly available, laboratory-confirmed weekly influenza case counts. The dataset spans six years, from January 2017 to December 2022, and includes a total of 313 weekly observations for Saudi Arabia. To simulate real-world forecasting conditions, the data were split chronologically into a training set (2017–2021; 260 weeks) and a hold-out test set (2022; 53 weeks).

### 2.2. Data Preprocessing and Feature Engineering

Data pre-processing was conducted to make the data complete, consistent, and suitable for training.

(a) Data cleaning: Inconsistent or incomplete weekly records were examined. Values above the permissible range (≈2% of total observations) were replaced through seasonal means interpolation in order to maintain the temporal connection.

(b) Temporal feature generation: Time features such as week of the year, month, and year were constructed to specifically represent seasonal and cyclical patterns of influenza incidence.

(c) Seasonal-trend decomposition: The long-term trends and periodic variances were separated using the STL algorithm, making it easier for the model to learn residual fluctuations [14, 16].

(d) Lag transformation and normalization: The data were reformulated in supervised learning with four lagged predictors (weeks t–1 to t–4) predicting incidence for week t.

Numerical values were z-score normalized, and min–max scaling was applied ([0, 1]) to the entire set of numerical variables in order to facilitate model convergence.

### 2.3. Model Development

#### *2.3.1. We Employed Two Machine Learning Regression Models*

<u>Support Vector Regression (SVR)</u>

Support Vector Regression (SVR) is a learning method based on the Support Vector Machine (SVM) formulation, which has been designed to predict continuous output [17,19]. Rather than being classified, the SVR algorithm tries to find an optimal regression function and the value of a target variable that is no more deviated from some margin ε. The trade-off in accessing the balance between an accurate model and generalization is done by minimizing the sum of errors with a penalty factor on deviations greater than ε. The regularization parameter C balances the trade-off between model complexity and fitting accuracy, and the kernel function transforms input samples to a higher-dimensional space in which non-linearity can be modeled [20,21].

In our study, we used an RBF (Radial Basis Function) kernel because it is quite flexible in modelling non-linear temporal and seasonal dependence of influenza incidence. The SVR model included lagged influenza cases and time terms (week, month, year) as predictors. Model hyperparameters (C, γ, and ε) of classifiers were adjusted using a grid search and a fivefold cross-validation strategy for the best generalization and fit to avoid over-fitting [22].

#### *2.3.2. Random Forest Regression (RF)*

Random Forest is an ensemble learning technique that generates a stronger and less general model by aggregating predictions from numerous decision trees [23]. Each tree in the ensemble is constructed by bootstrapping a sample of the training data. where the input features are randomized at each node split to reduce inter-tree correlation and thus enhance stability. The overall prediction is taken as the mean of all outputs, ensuring low variance and preventing overfitting in single tree models [24, 27].

Random Forest modelling was performed in this study by using the Scikit-learn toolkit for Python. The 2022 influenza circulation data was kept aside for the out-of-sample validation. Important hyperparameters, including the number of estimators (trees), maximum depth of trees, and minimum samples per leaf were tuned in a grid search manner using fivefold cross validation to minimize Root Mean Squared Error (RMSE). The feature importance ranking was based on the Mean Squared Error Reduction (MSER) criterion while random subspace sampling was used for obtaining a bias reduced model.

These settings enabled the RF model to learn nonlinear relationships and interactions between lagged influenza counts and temporal features. The ensemble form allows it to process the noise, seasonality and idiosyncratic variations present in real-world influenza surveillance data with good generalization performance for operational forecasting tasks in public health.

### 2.4. Model Training and Validation

The Python scikit-learn library was used to implement the models [28,31]. The dataset was divided into two parts by the year to maintain the time consistency *i.e.*, training on 2017-2021 data and testing on 2022 data. Training was performed using a five-fold cross-validation to reduce overfitting. The hyperparameters were optimized through grid search through the following ranges:

SVR: C ∈ {1, 10, 100}, kernel = RBF, γ ∈ {0.01, 0.1, 1}, ε ∈ {0.01, 0.1, 1}

RF: estimator's ∈ {100, 200, 300}, max_depth ∈ {5, 10, 20}, min_samples_leaf ∈ {1, 2, 4}

The best model configuration for each model was determined with the lowest RMSE across validation folds. Afterward, the models were left trained at 2017–2021 and fed with 2022 data to carry out the independent testing. This extensive validation procedure supports both reliability and generalizability of the forecasting results for national monitoring of influenza activity in Saudi Arabia.

#### *2.4.1. Performance Metrics*

Model performance was evaluated using four standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination ($R^2$). These provide a comprehensive assessment of prediction accuracy and error magnitude.

MAE: Measures the average magnitude of prediction errors, irrespective of direction:

$$\mathrm{MAE} = \frac{\sum_{j=1}^{n} |y_j - \hat{y}_i|}{n} \tag{1}$$

MSE: Computes the average squared difference between predicted and actual values.

RMSE: The square root of MSE, expressed in the same units as the target variable [32,35].

#### *2.4.2. Root Mean Square Error (RMSE)*

The subsequent Equation (5) expresses the Root Mean Square Error (RMSE) to assess the standard deviation of the forecast errors.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2}$$

#### *2.4.3. Coefficient of Determination ($R^2$)*

It is a statistical degree that calculates the amount of variance in the dependent variable (weekly influenza cases) that is expected from the independent variables. Its values range between 0 and 1. An $R^2$ value of 0 indicates that the model is unsuccessful to explicate any of the inconsistency in the response data dependent on the predictors, while a value of 1 specifies a perfect fit model, where all variance in the outcome variable is precisely reported for by the model.

The $R^2$ is premeditated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{\bar{y}}_i)^2} \quad (3)$$

Where;

$\hat{y}_i$ = Predicted value of the $i^{th}$ sample and represents the matching real value for the whole n samples [32].

### *2.4.4. Mean Absolute Percentage Error (MAPE)*

To provide an interpretable percentage-based error measure, the Mean Absolute Percentage Error (MAPE) was also computed as:

$$\text{MAPE} = \frac{100\%}{n}\sum_{j=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (4)$$

represent the observed and predicted values respectively.

### *2.4.5. Peak Week Error (PWE)*

In order to determine the accuracy of the peak timing prediction, which is important in epidemic preparedness, a Peak Week Error (PWE) was defined as:

$$\text{PWE} = |\ \text{Week}_{\text{predicted peak}} - \text{Week}_{\text{actual peak}}\ | \quad (5)$$

The metric measures the change (in weeks) between the predicted and observed maximum influenza incidence, a direct measurement of the temporal accuracy that is very pertinent in real-time preparedness to epidemics and population health response [36].

### 2.5. Ethical Considerations

The information utilized in this study is open and accessible in the databases of WHO. There were no human subjects used and it did not need any ethical consent.

## 3. RESULTS

### 3.1. Descriptive Statistics

Weekly influenza cases ranged from 7 to 546 (mean = 157.9, SD = 116.3). The STL decomposition showed frequent winter peaks and a marine trend of an increase in intensity over the six-year extension (Figure **1**). This nonlinear time-series forecasting model is justified by this seasonal trend.

Cases per week ranged between 7 and 546, with a mean of 157.9 and a standard deviation of 116.31. Figure **1** shows the seasonal-trend decomposition that is performed with the help of STL. The decomposition verified a regular seasonal pattern with regular peaks in the winter season. Another trend that observed was a slight upward trend in peak intensity over the six years period implying that there has been a slow growth in the magnitude of these epidemics over the years. Such seasonal dynamics facilitate the use of non-linear learning-based time series forecasting models.

### 3.2. Model Performance

SVR and RF models were trained on 2017–2021 data and tested on 2022 data. Results are summarized in Table **1**.

Although SVR achieved superior fit during training, its performance deteriorated on the test set, with higher prediction error and lower $R^2$. In contrast, RF maintained better generalization, exhibiting lower RMSE and higher $R^2$ on 2022 data—indicating greater robustness to unseen seasonal dynamics.

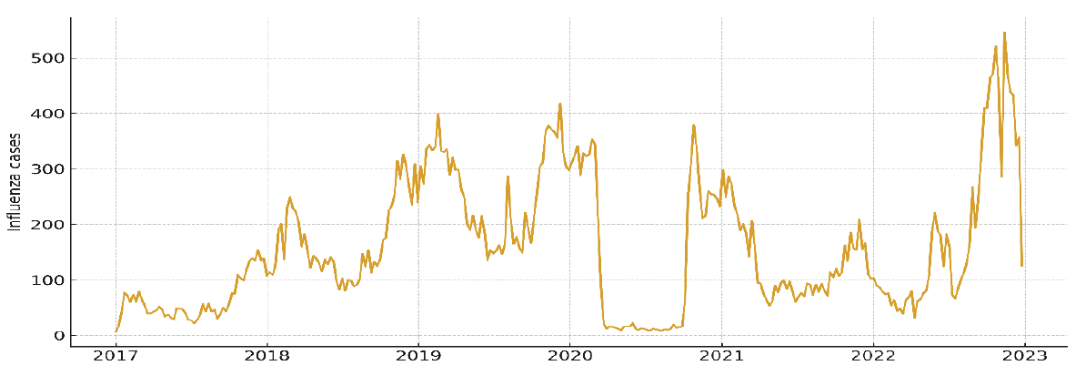


**Figure 1:** Weekly influenza cases in Saudi Arabia (2017–2022), showing winter peaks and increasing epidemic intensity via STL decomposition.

**Table 1: Model Training and Test Performance Metrics**

| Model | Dataset | Training | | | |
|---|---|---|---|---|---|
| | | MAE | MSE | RMSE | R² |
| Support Vector Regression | Training | 2.85 | 13.485 | 3.672 | 0.960 |
| | Test | 51.786 | 5739.090 | 75.757 | 0.750 |
| Random Forest | Training | 13.485 | 408.930 | 20.222 | 0.891 |
| | Test | 41.590 | 51.78 | 7.196 | 0.818 |

### 3.3. Visual Prediction Analysis

Figure **2** shows that SVR followed the general trend but significantly underestimated peak values during high-incidence periods, lagging behind sharp spikes. In contrast, Figure **3** demonstrates that RF aligned closely with observed case counts, especially during epidemic peaks. The ensemble nature of RF enabled it to capture sharp, nonlinear variations more effectively.

### 3.4. Residual Analysis

Figure **4** supports the sensitivity of SVR to fast surges but RF shows constant residual values, both in amplitude and in the timing of outbreaks. The high value of RF in epidemic preparedness is emphasized by its high quality of real-time predictions. In Saudi Arabia, where mass events and the risk of spreading respiratory infections are high, the possibility to forecast the weekly influenza activity properly is an essential factor in regard to the timely intervention. RF was also reliable in trend detection and amplitude estimation, which reinforced its possible functionality in the framework of the public health surveillance to maximize the allocation of vaccines, decision-making, and responses to early-warnings.

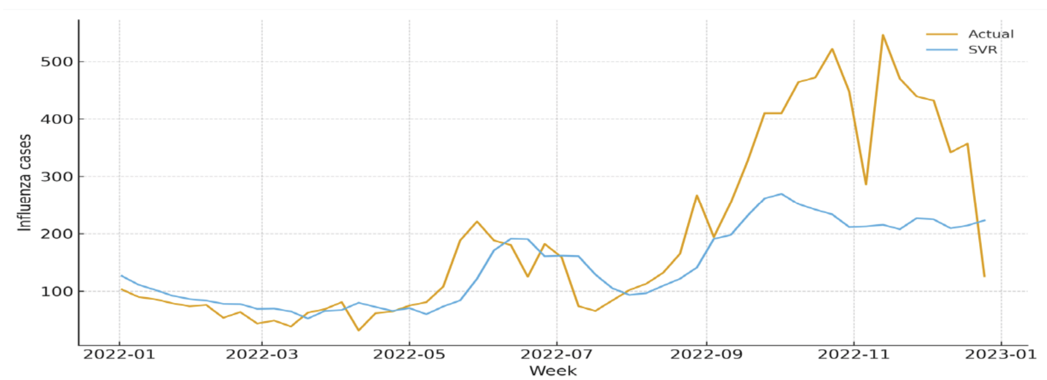


**Figure 2:** SVR-predicted vs. observed influenza cases (2022); captures pattern but underestimates peak magnitudes.

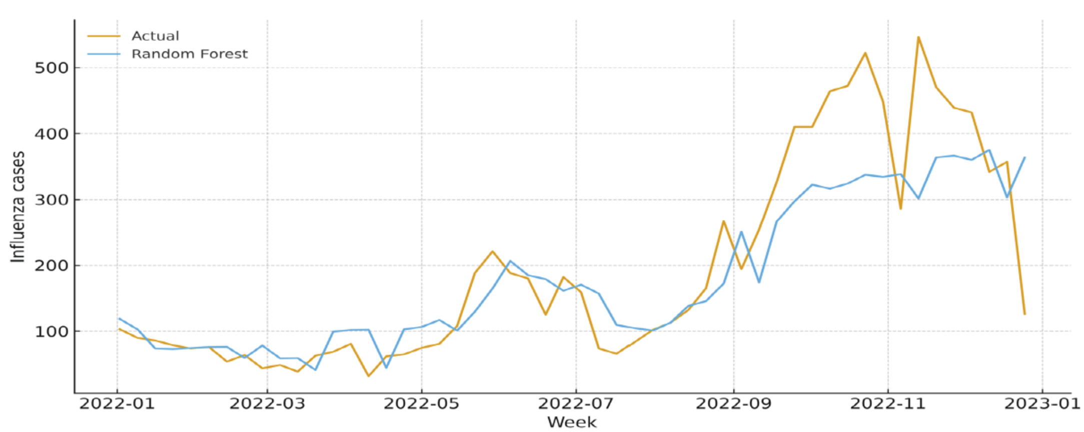


**Figure 3:** RF-predicted vs. observed influenza cases (2022); accurately reproduces timing and amplitude of peaks.

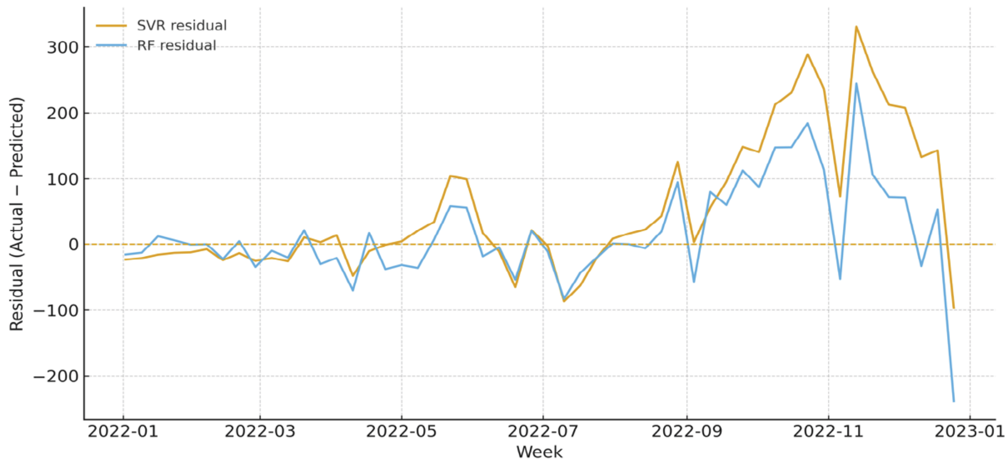


**Figure 4:** Weekly residual errors (SVR vs. RF, 2022); RF shows stable, near-zero residuals compared to SVR's wider fluctuations.

### 3.5. Uncertainty Quantification

Table **2** presents representative RF forecasts with 95% prediction intervals, illustrating its reliability across seasonal fluctuations.

**Table 2: Representative Weekly RF Forecasts with 95% Prediction Intervals**

| Date | Actual | Predicted | Lower 95% | Upper 95% |
|---|---|---|---|---|
| 2-Jan-22 | 103 | 130.2 | 72 | 286 |
| 13-Feb-22 | 54 | 62.4 | 22 | 95 |
| 04-Sep-22 | 194 | 244.3 | 201 | 315 |
| 13-Nov-22 | 546 | 284.0 | 207 | 368 |
| 25-Dec-22 | 126 | 381.8 | 289 | 418 |

### 3.6. Seasonal Error Distribution

**Table 3: Seasonal Model Error Comparison (2022)**

| Season | Model | MAE | RMSE | $R^2$ |
|---|---|---|---|---|
| Peak (Oct–Mar) | SVR | 99.09 | 141.48 | 0.42 |
| Peak (Oct–Mar) | RF | 64.79 | 96.43 | 0.73 |
| Trough (Apr–Sep) | SVR | 43.81 | 58.12 | 0.57 |
| Trough (Apr–Sep) | RF | 38.46 | 48.31 | 0.70 |

RF demonstrated greater stability during both peak and trough phases.

## 4. DISCUSSION

This study advances regional influenza forecasting by delivering a validated, multi-year ML framework tailored to Saudi Arabia's surveillance context. Unlike prior applications that often rely on short time spans or proxy signals, we leverage six years of laboratory-confirmed national data and evaluate models under a temporal hold-out that mimics real-world deployment. Methodologically, we demonstrate that an ensemble approach (RF) provides stronger generalization to unseen seasons and greater resilience to nonlinear surges than a kernel-based baseline (SVR). Substantively, the framework is region-aware, capturing winter seasonality and accommodating mass-gathering effects (e.g., Hajj/Umrah) through a lagged/seasonal structure, while remaining robust despite routine noise in the surveillance series.

i. Relative to prior work from East Asia and other settings, our contribution is threefold:

ii. Design for operational use: a train/validate/test split aligned with surveillance workflows and prospective evaluation;

iii. Actionable uncertainty: pairing point forecasts with 95% prediction intervals and Peak-Week Error for preparedness timing;

iv. Policy-facing interpretation: mapping model outputs to public-health triggers (e.g., surge detection and vaccine logistics) under local constraints.

These advances position the model as a deployable early-warning component rather than a retrospective benchmarking exercise.

## 5. FROM MODEL TO PUBLIC HEALTH ACTION

We advocate for a straightforward and unambiguous route that directly links influenza predictions to public health interventions.

The Random Forest algorithm can produce weekly forecasts with 95% prediction intervals utilizing automated data updates from the WHO. These projections can directly inform public health dashboards that present real-time and short-term predictions, facilitating effective planning. Upon identifying a probable rise, signaled by projected instances exceeding the historical upper limit, the model must automatically alert public health officials. This will help them improve surveillance, testing, ensure there are enough vaccines and antivirals, and prepare hospitals for an influx of patients. Consistent retraining and performance evaluations will ensure that the forecasting system remains dependable, transparent, and aligned with national health priorities.

## 6. LIMITATIONS

This research possesses multiple limitations. The research utilized national surveillance data, which may not accurately reflect local disparities or reporting delays. Climatic and demographic influences on influenza spread were not considered, and the models may still be susceptible to some mild overfitting, despite cross-validation. Prospective investigations should consider incorporating meteorological, population, and behavioral data to refine risk prediction with greater regional generalizability.

## 7. IMPLICATIONS AND FUTURE DIRECTIONS

### 7.1. Future Enhancements may Include

Machine learning, fractional calculus, stochastic simulation and optimal control are promising directions for an integrated intelligent system of flu forecasting. By integrating data-driven models with mathematically based strategies, public health authorities can develop adaptive systems that not only update predictions and quantify the degree of uncertainty but also adjust measures in real-time as epidemic patterns emerge.

Recent advances in fractional and stochastic modeling have shown that these approaches more accurately capture the memory, randomness, and behavioral variability inherent in infectious disease transmission [45-47]. At the same time, optimal control theory provides a structured foundation for evaluating and fine-tuning intervention policies—such as vaccination, antiviral treatment, and public awareness programs—under practical constraints of time and resources [48, 49].

When used together, these approaches have the potential to move influenza modeling from being merely predictive and passive, towards an active and data-informed decision-support tool. This transition shifts public health action to an epidemic preparedness approach, which is consistent with Saudi Arabia's broader vision for anticipatory health preparedness and evidence-based responses to new infectious perils.

## 8. CONCLUSION

The study confirms that RF outperforms SVR in forecasting weekly influenza trends in Saudi Arabia, combining accuracy, generalization, and practical interpretability. The framework is reproducible, scalable, and adaptable for integration into national surveillance systems. As the global focus shifts toward predictive public health, such ML-based systems will play an essential role in epidemic preparedness and resource optimization.

## ETHICAL APPROVAL AND DATA AVAILABILITY

All data are publicly available from the WHO GISRS platform: https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal). No human participants were involved; thus, ethical approval was not required.

## ACKNOWLEDGEMENTS

The authors thank the World Health Organization (WHO) for maintaining open-access influenza surveillance data.

## DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest.

## FUNDING

This research received no specific grant from public, commercial, or not-for-profit funding agencies.

## DATA AVAILABILITY STATEMENT

Publicly available from WHO GISRS

## REFERENCE

[1] World Health Organization. (2023). Influenza (seasonal). https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

[2] Alshamrani M, Farahat F, Alzunitan M, *et al*. Hajj vaccination strategies: Preparedness for risk mitigation. J Infect Public Health 2024; 17(11): 102547. https://doi.org/10.1016/j.jiph.2024.102547

[3] Maddah N, Verma A, Almashmoum M, *et al*. Effectiveness of public health digital surveillance systems for infectious disease prevention and control at mass gatherings: systematic review. J Med Internet Res 2023; 25: e44649. https://doi.org/10.2196/44649

[4] Guma FE. Analysis of influenza-like illness trends in Saudi Arabia: a comparative study of statistical and deep learning techniques. Osong Public Health Res Perspect 2025; 16(3): 270-284. https://doi.org/10.24171/j.phrp.2025.0080

[5] Alzahrani SM, Guma FE. Improving seasonal influenza forecasting using time series machine learning techniques. J Inf Syst Eng Manage 2024; 9(4): 30195. https://doi.org/10.55267/iadt.07.15132

[6] Daqqa I, Almarashi AM, Bashier MM, Aripov M, Abaker AOI, Alhag AA, Shokeralla AA. Predictive modeling of breast cancer incidence: A comparative study of fuzzy time series and machine learning techniques. Journal of Statistics Applications & Probability 2024; 14(2): 183-189. https://doi.org/10.18576/jsap/140204

[7] Rodriguez A, Kamarthi H, Agarwal P, *et al*. Machine learning for data-centric epidemic forecasting. Nat Mach Intell 2024; 6(10): 1122-1131. https://doi.org/10.1038/s42256-024-00895-7

[8] Keshavamurthy R, Dixon S, Pazdernik KT, *et al*. Predicting infectious disease for biopreparedness and response: a systematic review of machine learning and deep learning approaches. One Health 2022; 15: 100439. https://doi.org/10.1016/j.onehlt.2022.100439

[9] Badami V. A study on influenza reports: the impact of various influencing factors and a predictive modeling approach to forecasting flu cases in European countries [doctoral dissertation]. Dublin, Ireland: National College of Ireland; 2022.

[10] Santangelo OE, Gentile V, Pizzo S, Giordano D, Cedrone F. Machine learning and prediction of infectious diseases: a systematic review. Mach Learn Knowl Extr 2023; 5(1): 175-198. https://doi.org/10.3390/make5010013

[11] Babanejaddehaki G, An A, Papagelis M. Disease outbreak detection and forecasting: a review of methods and data sources. ACM Trans Comput Healthc 2025; 6(2): 1-40. https://doi.org/10.1145/3708549

[12] Martin-Moreno JM, Alegre-Martinez A, Martin-Gorgojo V, *et al*. Predictive models for forecasting public health scenarios: practical experiences applied during the first wave of the COVID-19 pandemic. Int J Environ Res Public Health 2022; 19(9): 5546. https://doi.org/10.3390/ijerph19095546

[13] Shokeralla AA, Alzharani AA, Abdalla AH, Modawy YM, Elshamy I, El Guma F. Modeling Climate-Driven Cholera Outbreaks: A Negative Binomial Regression Framework with Improved Handling of Overdispersion and Extreme Events. Letters in Biomathematics 2025; 12(1).

[14] Zhang AR, Li XL, Wang T, *et al*. Ecology of Middle East respiratory syndrome coronavirus, 2012–2020: A machine learning modelling analysis. Transbound Emerg Dis 2022; 69(5): e2122-e2131. https://doi.org/10.1111/tbed.14548

[15] Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: A seasonal-trend decomposition. J Off Stat 1990; 6: 3-73.

[16] Vapnik VN. The Nature of Statistical Learning Theory. New York, NY: Springer 1995. https://doi.org/10.1007/978-1-4757-2440-0

[17] Sharma S, Raj K, Sharma SK, Khalid TA, Mustafa AO, Mohammed MM, Bakery AA. Applications of strongly deferred weighted convergence in the environment of uncertainty. International Journal of Analysis and Applications 2024; 22: 181-181. https://doi.org/10.28924/2291-8639-22-2024-181

[18] Pedregosa G, Varoquaux A, Gramfort A, *et al*. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011; 12: 2825-2830.

[19] Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004; 14(3): 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

[20] Sykes ER, Wang Y. Optimising SVR for epidemiological predictions: a case study on COVID-19 mortality in Japan. Int J Artif Intell Soft Comput 2024; 8(5): 1-29. https://doi.org/10.1504/IJAISC.2024.143383

[21] Shokeralla AA. The Discrete Laplace Transform (DLT) Order: A Sensitive Approach to Comparing Discrete Residual Life Distributions with Applications to Queueing Systems. European Journal of Pure and Applied Mathematics 2025; 18(4): 6994. https://doi.org/10.29020/nybg.ejpam.v18i4.6994

[22] Zhang J, Nawata K. A comparative study on predicting influenza outbreaks. Biosci Trends 2017; 11(5): 533-541. https://doi.org/10.5582/bst.2017.01257

[23] Zhao Z, Zhai M, Li G, *et al*. 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 2023; 23(1): 71. https://doi.org/10.1186/s12879-023-08025-1

[24] Shokeralla AA, Qurashi ME, Mekki RY, Ali MS. The Effect of Symptoms on the Survival Time of Coronavirus Patients in the Sudanese Population. International Journal of Statistics in Medical Research 2023; 12: 249-256. https://doi.org/10.6000/1929-6029.2023.12.29

[25] Al-Tawfiq JA, Gautret P, Benkouiten S, Memish ZA. Mass gatherings and the spread of respiratory infections. Lessons from the Hajj. Ann Am Thorac Soc 2016; 13(6): 759-765. https://doi.org/10.1513/AnnalsATS.201511-772FR

[26] Memish ZA, Zumla A, Alhakeem RF, *et al*. Hajj: infectious disease surveillance and control. Lancet 2014; 383(9934): 2073-2082. https://doi.org/10.1016/S0140-6736(14)60381-0

[27] Muhammad SK, Ansari TA, Shabbir B, *et al*. The role of artificial intelligence in public health surveillance: a post-pandemic perspective. Insights J Life Soc Sci 2025; 3(3): 74-80. https://doi.org/10.71000/pa7ab080

[28] Saadeh R, Shokeralla AA, Al-Kuleab N, Hamad WS, Ali M, Abdoon MA, El Guma F. Stochastic Modelling of Seasonal Influenza Dynamics: Integrating Random Perturbations and Behavioural Factors. European Journal of Pure and Applied Mathematics 2025; 18(3): 6379-6379. https://doi.org/10.29020/nybg.ejpam.v18i3.6379

[29] Khalid TA, Imam A, Bahatheg A, Elsamani SA, El Mukhtar B. A fractional epidemiological model for prediction and simulation the outbreaks of dengue fever outbreaks in sudan. Journal of Survey in Fisheries Sciences 2023; 10(3S): 2679-2692.

[30] Chen Q, Zheng X, Shi H, *et al*. Prediction of influenza outbreaks in Fuzhou, China: comparative analysis of forecasting models. BMC Public Health 2024; 24(1): 1399. https://doi.org/10.1186/s12889-024-18583-x

[31] Bezerra AKL, Santos ÉMC. Prediction the Daily Number of Confirmed Cases of COVID-19 in Sudan with ARIMA and Holt Winter Exponential Smoothing. International Journal of Development Research 2020; 10(08): 39408-39413.

[32] Cong J, Ren M, Xie S, Wang P. Predicting seasonal influenza based on SARIMA model, in mainland China from 2005 to 2018. Int J Environ Res Public Health 2019; 16(23): 4760. https://doi.org/10.3390/ijerph16234760

[33] Shokeralla AA. A Hybrid Time Series–Regression Model for Tuberculosis Forecasting in Resource-Limited Settings, International Journal of Statistics in Medical Research 2025; 14: 299-307. https://doi.org/10.6000/1929-6029.2025.14.29

[34] Babanejaddehaki G, An A, Papagelis M. Disease outbreak detection and forecasting: a review of methods and data sources. ACM Trans Comput Healthc 2025; 6(2): 1-40. https://doi.org/10.1145/3708549

[35] Montesinos López OA, Montesinos López A, Crossa J. Support vector machines and support vector regression. In Multivariate statistical machine learning methods for genomic prediction. Cham: Springer International Publishing 2022; pp. 337-378. https://doi.org/10.1007/978-3-030-89010-0_9

[36] Kakarash ZA, Ezat HS, Omar SA, Ahmed NF. Time series forecasting based on support vector machine using particle swarm optimization. International Journal of Computing 2022; 21(1): 76-88. https://doi.org/10.47839/ijc.21.1.2520

[37] Szostek K, Mazur D, Drałus G, Kusznier J. Analysis of the Effectiveness of ARIMA, SARIMA, and SVR Models in Time Series Forecasting: A Case Study of Wind Farm Energy Production. Energies (19961073) 2024; 17(19). https://doi.org/10.3390/en17194803

[38] del-Pozo-Bueno D, Kepaptsoglou D, Peiró F, Estradé S. Comparative of machine learning classification strategies for electron energy loss spectroscopy: Support vector machines and artificial neural networks. Ultramicroscopy 2023; 253: 113828. https://doi.org/10.1016/j.ultramic.2023.113828

[39] Bountzis P, Kavallieros D, Tsikrika T, Vrochidis S, Kompatsiaris IY. A deep one-class classifier for network anomaly detection using autoencoders and one-class support vector machines. Frontiers in Computer Science 2025; 7: 1646679. https://doi.org/10.3389/fcomp.2025.1646679

[40] Mustaffa Z, Sulaiman MH. Random forest-based wind power prediction method for sustainable energy system. Cleaner Energy Systems 2025; 100210. https://doi.org/10.1016/j.cles.2025.100210

[41] Tjøstheim D. Selected topics in time series forecasting: Statistical models vs. machine learning. Entropy 2025; 27(3): 279. https://doi.org/10.3390/e27030279

[42] Ibrahim A, Kashef R, Corrigan L. Predicting market movement direction for bitcoin: A comparison of time series modeling methods. Computers & Electrical Engineering 2021; 89: 106905. https://doi.org/10.1016/j.compeleceng.2020.106905

[43] Gajewski P, Čule B, Rankovic N. Unveiling the power of ARIMA, support vector and random forest regressors for the future of the Dutch employment market. Journal of Theoretical and Applied Electronic Commerce Research 2023; 18(3): 1365-1403. https://doi.org/10.3390/jtaer18030069

[44] Magalhães B, Bento P, Pombo J, Calado MDR, Mariano S. Short-term load forecasting based on optimized random forest and optimal feature selection. Energies 2024; 17(8): 1926. https://doi.org/10.3390/en17081926

[45] Pedregosa G, Varoquaux A, Gramfort A, *et al*. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011; 12: 2825–2830.

[46] Zhu X, Ren J, Wang J, Li J. Automated machine learning with dynamic ensemble selection. Appl Intell 2023; 53(20): 23596-23612. https://doi.org/10.1007/s10489-023-04770-7

[47] Dixon S, Keshavamurthy R, Farber DH, *et al*. A comparison of infectious disease forecasting methods across locations, diseases, and time. Pathogens 2022; 11(2): 185. https://doi.org/10.3390/pathogens11020185

[48] Guma FE, Abdoon MA, Abdalla SJM, Alharbi SA, Alsemiry RD, Allogmany R, Al-Kuleab N. Fractional Order Modelling of Influenza Dynamics: Impact of Memory Effect and Severity-Based Variations. International Journal of Biomathematics 2025.

[49] Ali M, Guma FE, Qazza A, Saadeh R, Alsubaie NE, Althubyani M, Abdoon MA. Stochastic modeling of influenza transmission: Insights into disease dynamics and epidemic management. Partial Differential Equations in Applied Mathematics 2024; 11: 100886. https://doi.org/10.1016/j.padiff.2024.100886

Received on 02-10-2025 Accepted on 05-11-2025 Published on 24-11-2025

https://doi.org/10.6000/1929-6029.2025.14.64