

Research Article

Using Sequence Mining to Predict Complex Systems: A Case Study in Influenza Epidemics

Theyazn H. H. Aldhyani ¹, **Manish R. Joshi** ², **Shahab A. AlMaaytah**,¹
Ahmed Abdullah Alqarni ³ and **Nizar Alsharif**⁴

¹Community College of Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa, Saudi Arabia

²School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, MS, India

³Department of Computer Sciences and Information Technology, Albaha University, Al Bahah, Saudi Arabia

⁴Department of Computer Engineering and Science, Albaha University, Al Bahah, Saudi Arabia

Correspondence should be addressed to Theyazn H. H. Aldhyani; taldhyani@kfu.edu.sa

Received 18 March 2021; Revised 29 March 2021; Accepted 2 April 2021; Published 15 April 2021

Academic Editor: M. Irfan Uddin

Copyright © 2021 Theyazn H. H. Aldhyani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

According to the World Health Organisation, three to five million individuals are infected by influenza, and around 250,000 to 500,000 people die of this infectious disease worldwide. Influenza epidemics pose a serious public health threat. Moreover, graver dangers are encountered with influenza subtypes against which there is little or no preexisting human immunity. Such subtypes of influenza have the potential to cause devastating epidemics. Thus, enhancing surveillance systems for the purpose of detecting influenza epidemics in an early stage can quicken response times and save millions of lives. This paper presents three adapting intelligence models: support vector machine regression (SVMR), artificial neural network using particle swarm optimisation (ANNPSO), and our intelligent time series (INTS) to predict influenza epidemics. The novelty of the current study is that it proposes a new intelligent model to predict influenza outbreaks. The INTS model combines clustering with a time series model to enhance the prediction of influenza outbreaks. The innovation of our proposed model integrates the results obtained from the existing weighted exponential smoothing model with centroids obtained from clustering. We developed a surveillance system for influenza epidemics using Google search queries. The current research is based on a weighted version of the Center for Disease Control and Prevention influenza-like illness activity level obtained from the Center for Disease Control and Prevention data, as well as query data obtained from the Goggle search engine in the USA. The influenza-like illness data was collected from January 4, 2009 (week 1), to December 27, 2015 (week 52), stretching across a total time span of 312 weeks. Google Correlate was used to select search queries related to influenza epidemics. In total, 100 search queries were obtained from Google Correlate, 10 of which were better and more relevant search queries selected in this study. The model was evaluated using online Google search queries collected from Google Correlate. Standard measure performance MSE, RMSE, and MAE were employed to estimate the results of the proposed model. The empirical results of the INTS model showed MSE = 0.003, RMSE = 0.036, and MAE = 0.0185, indicating that the errors of the proposed model are very limited. A comparative model of predicting results between the INTS model, alternative Google Flu Trend (GFT), and autoregression with Google search data is also presented. The proposed model outperformed the existing models.

1. Introduction

With the rapid development of societies and economies worldwide, health technologies have been enhanced, and health facilities have been promoted as well. The flu infection faces societies with a number of health problems.

Consequently, influenza diseases have still posed a great threat to human health, and controlling influenza diseases has become a very important challenge globally. Influenza has brought huge losses to national economies and continues to pose a serious threat to human health across the world. Although the subtypes of influenza diseases, such as

smallpox and malaria, have been efficiently controlled, the seasonal incidences of influenza still have high occurrence rates and cause many emergent health problems, including early deaths worldwide [1].

Therefore, influenza is the first infectious disease for which a surveillance system was implemented. Yet, its effective control remains elusive. Millions of Internet users around the world have submitted Internet search terms for the purpose of developing a system to detect influenza outbreaks at the earliest stages [2, 3]. The rapid adoption of the Internet has opened new gates for developing and enhancing healthcare. Many researchers have used huge amounts of data on the Internet and social media platforms such as Twitter or Facebook to discover novel methods to diagnose diseases. Thus, the language patterns from the Internet and social media have proved their usefulness in analysing and predicting chronic diseases and in determining the behaviours and habits that increase the possibility of those diseases. Understanding population behaviour and trends of noncommunication diseases is directed by using web search activity data. Noncommunicable diseases have been detected by using web search activity data and examination data that has been submitted to the concerned health officials. These search activity data have the same trend as examination data [4, 5].

Researchers have compared Internet search query data relating to the main key adaptable risk factors of non-communicable diseases with clinical population data from the US Center for Disease Control. Developing real-time surveillance can provide a proxy for clinical population data and real-time web search data for enhancing healthcare systems. Most previous research has tried to predict influenza disease using data from the Internet search query alone. Here, we developed a new model that has the capability to predict the influenza epidemic with the best accuracy.

The main contribution of this study is to propose a complex system that can assist in enhancing the time series models in the healthcare domain. The INtelligent Time Series (INTS) combines clustering with a time series mode. INTS model was developed to predict the influenza epidemic based on a Google search. We found that the INTS model is capable of yielding better results compared with another proposed model, such as Google Flu Trend (GFT) and Auto-Regression with Google search data (ARGO).

2. Background of the Study

Typically, the Internet is a primary tool that can identify individuals making attempts towards wellbeing and supplying data. Individuals are frequently subjected to certain infections or medicinal problems and always look for suitable medicinal medications or methods. Various studies have recommended remarkable methods for predicting influenza epidemics [6, 7]. In November 2008, Google launched the Google Flu service, which uses a computational search term model to predict influenza activity. In 2009, Google also offered Google Flu Trends (GFT), a digital method used to detect public health surveillance [8]. By gathering web information, the investigator claims to validly

estimate influenza epidemics. The novelty of the GFT model is that it is used by the Center for Disease Control (CDC) to find specific search terms from digital data for predicting influenza epidemics. Various subsequent studies have modelled their approaches after the GFT model to enhance the GFT model [9–12].

Hence, we present the INtelligent Time Series (INTS) model, which outperforms all alternative models in predicting influenza epidemics by using Internet search queries. Increasing studies are focusing on monitoring data-based infectious illnesses to complement current technologies and develop new models [13–19]. Furthermore, developed models for detecting infectious diseases using Internet searches are presently being conducted by using large amounts of information, such as Internet search queries [20–24]. Thus, it becomes possible to collect and process Internet search information to monitor the healthcare system. Internet search information has the ability to detect an epidemic at a better speed than standard surveillance technologies, according to Towers et al. [25]. For instance, the model that included search query data obtained the best results when Huang et al. predicted hand, foot, and mouth disease using the generalized additive model (GAM). As such, fresh big data surveillance tools have been shown to have the benefit of easy accessibility and recognising patterns in infectious disease before formal organisations [26]. Social media provides big data for useful information that can help discover those patterns. Tenkanen et al. reported that big data on social media is comparatively simple to obtain useful information for developing a real-time system [27]. This proposed research uses Twitter information to forecast mental illness [28]. Besides the influenza epidemic, a new type of influenza virus against which there is no previous immunity shows human-to-human transmission and has caused millions of deaths until an epidemic vaccine has been discovered. This system can use search queries from Google's search engine for influenza epidemic surveillance [29]. Quick and early estimation and prediction of the influenza epidemic before spreading greatly helps governments, health officials, and healthcare organisations to take appropriate decisions and timely prevention measures. In addition, influenza epidemic surveillance helps to provide information about the spread of influenza on a larger scale. Furthermore, the system helps in taking preemptive measures and spreading awareness regarding the disease to minimize its spread. The increased number of Internet users and researchers has helped identify Google's search engine use as a new monitoring scheme to complement the traditional scheme. Thus, Google Flu Trend tracks Google queries for obtaining information linked to influenza behaviour by Google customers, which shows a correlation with influenza CDC data while providing a projection of 1 to 2 weeks before CDC releases. Researchers and developers have presented several techniques to accomplish real-time surveillance systems for controlling the spread of influenza. It has also been demonstrated that the attention methodology for widely enhancing the GFT model with digital disease detection shows a guaranteed value. The attention model chooses particular queries automatically to monitor the

updated influenza epidemic estimate system. Mauricio et al. [30] collected Internet search terms from a clinician's database to forecast influenza activity. Santillana et al. [12] demonstrated how an alternative methodology enhancing the GFT model with guaranteed value for digital disease detection is broadly employed. The alternative model automatically chooses exact query terms to monitor the updates of the proposed model for estimating the influenza epidemic. Kang et al. [31] and Milinovich et al. [32] have developed a model to predict influenza activity by using Internet search queries obtained from influenza surveillance facilities in China. Milinovich et al. [33] presented a framework to estimate infectious diseases in Australia by using Internet search queries. They observed that web search activities have a potential role in predicting emerging infectious disease events. Samantha Cook et al. [34] compared their methodology to the GFT model for the estimation of influenza incidences by using an Internet search. They noticed that the model performed better than the GFT model. Nsoesie et al. [35] and Chretien et al. [36] presented a useful literature review of work in this area and described the methodology and data that estimate and predict the influenza epidemic. As they pointed out, some researchers used search queries to forecast influenza outbreaks. Twitter, Facebook, and Four Square are examples of sites where individuals intentionally post updates on their daily behaviours, health status, and physical locations. Paul et al. [37] used search queries from the social medium of Twitter to improve influenza forecasting. They observed that tweets were positively correlated with existing surveillance data provided by the CDC. HarshavardhanAchreka et al. [38] developed digital flu surveillance using Twitter data to estimate and predict the influenza epidemic. They argued that tweets collected from the social medium of Twitter could substantially help to detect influenza outbreaks earlier. Aldhyani et al. [39] proposed the adaptive network fuzzy inference system (ANFIS) model to predict chronic diseases using Google trend data. Using advanced artificial intelligence to diagnose diseases [40–42].

Thus, the objective of the present research was to build a model that assists in predicting the influenza epidemic using Google search queries. We integrated machine intelligence with the existing time series model to enhance the prediction of the influenza epidemic.

3. Materials and Methods

3.1. Data Sets

3.1.1. Epidemiological Surveillance Data. A weighted version of CDC'S influenza-like illness (ILI) activity level data was obtained from the Center for Disease Control and Prevention, which routinely collected epidemiological data and national statistics about influenza incidences on a weekly basis. We collected the ILI data from January 4, 2009 (week 1) to December 27, 2015 (week 52), across a total period of 312 weeks. This period covered the data expressed during the influenza seasons from the CDC in the USA. Data from the CDC ILINet system were obtained from [43], which

provides weekly influenza surveillance information at the national and regional levels of outpatient and viral illnesses. We decided to use the CDC ILI because the CDC data is a very strong data set. All reports about the CDC ILI are made available [44].

3.1.2. Google Correlate. Google search engines have become a significant part of everyone's lifestyle. They have become an indivisible clue for understating our lives. The Google search engine helps us search for an individual or an area and provides us with important information about events, problems, solutions, and other stuff. Many search engines are available, such as Google, Bing, AOL, Yahoo, and the like. Since Google is the most famous search engine, we searched for models using Google's centre. Google has a Google Trends centre that provides statistics on search queries conducted around the globe, place, and moment. Google has the facility of Google Trends, which provides the statistics of the searches made in the world with respect to the search query, location, and time. The Google Flu Trends is a good example of this use to predict the influenza epidemic. Our ultimate objective is to construct a model similar to the GFT system and other standard designs using open-source information and enhanced methodology. Our objective in information collection is to discover open-source search query information that looks like search queries used for GFT. For each season, ILI information could be acquired from the CDC as our basis. GFT system exploited 50 million of the most popular database queries in the United States, where a request was described as a full series of customer words, to discover some queries mostly linked to CDC information. Using a simpler technique, Google also constructed Google Correlate, which would provide information that a customer could upload and the corresponding daily time series of these queries, with the top 100 most associated search queries at the domestic stage. Therefore, it has used this tool to obtain an open-source dataset that reasonably matches the query data used in the GFT model that would not be released by Google. From January 4, 2009 (week 1) to December 27, 2015, we posted weighted CDC ILI information to Google Correlate and achieved an output of the 100 most important database queries. For each of these Correlate queries, the time series was not the real amount, but the quantity was subtracted by the median and split by the standard deviation of the sample. The output time series also ranged from January 4, 2004, through January 24, 2016. The output time series also ranged from January 4, 2004, through Jan 24, 2016. The Google Correlate has standardized the search volume of each query to have means zero and standard deviation one across time and contains data only from 2004 to January 2016. We compared our model with the original and revised (October 2014) Google flu Trend models. We observed that all the search terms obtained from Google Correlate were related to influenza activity. The 10 search terms that have heights correlated are selected search terms for predicting the influenza epidemic in this work.

To make Google Correlate data compatible with trend data, the min-max normalisation method was used. The

min-max method is used for scaling the data between 0 and 2. Table 1 shows the entire spurious search, which contains information about the influenza epidemic. We obtained 100 search queries, the same as the GFT model, but we selected only 10 search queries on sum each week, respectively. Table 2 displays the 10 Google search terms selected by our model. In comparison with GFT, ARGO, and other models, the GFT model selected 45 out of 100 terms each week on average, and the ARGO model selected 14 out of 100 terms each week on average, respectively. Google Correlate is available [38].

3.2. Normalisation Method. Normalisation transformation of the appropriate time series typically helps to improve suitable time series models. The Min-max method is employed in Matlab for scaling the data. This method transforms data within a range of 0 to 2 scales:

$$z_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (\text{New}_{\max_x} - \text{New}_{\min_x}) + \text{New}_{\min_x}, \quad (1)$$

where x_{\min} is the minimum of data and x_{\max} maximum of data. New_{\min_x} is minimum number 0 and New_{\max_x} is maximum number 2.

3.3. Prediction Models. In this section, the proposed system is presented. Figure 1 shows the generic framework of the proposed system.

3.3.1. The Intelligent Time Series (INTS) Model. The INTS model explicitly predicts influenza outbreaks using Google search queries. Figure 2 illustrates how the INTS model can be a hybrid model with the existing time series prediction model and the k -means clustering algorithm. The prediction model was used to predict the influenza epidemic using Google search queries. Furthermore, the k -means clustering algorithm is employed to analyze the search pattern that has been obtained from the Google engine separately. The novelty of the INTS model lies in its integration of the results obtained from the WES time series prediction model along with the centroid obtained from k -means algorithms. The INTS model is a function of results obtained from the WES model and centroids of the k -means clustering algorithm.

$EP_i = f(P_i, C_i)$ is prediction function generated from the time series model and the centroid of clustering. The integrating model improved the prediction results. A comparative prediction result between the INTS model and existing times series models is presented. It is noted that the INTS model outperforms. The steps of the proposed INTS algorithm are discussed in the following subsections. The INTS algorithm is shown below: Let, S_i be the sample of i^{th} day, K be the number of clusters and K_i be the i^{th} cluster, C_i is the centroid of i^{th} cluster. Let P_i be the prediction for i^{th} sample obtained by using WES model and EP_i is an enhanced prediction for the i^{th} sample obtained by using the proposed model (Algorithm 1).

The components of the INTS proposed system are as follows:

(1) *Weighted Exponential Smoothing (WES) Model.* Exponential smoothing models are one of the most important prediction approaches widely used in industry and commerce.

$$\ell_0 = \bar{y} = \frac{\sum_{t=1}^n y_t}{n}, \quad (2)$$

$$S_{T+1} = \alpha y_t + (1 - \alpha)S_t,$$

where S_{T+1} remains constant and is smoothing data.

The exponential smoothing method is the generalisation of the moving average technique. Exponential smoothing models are also one of the prediction approaches that use stationary time series data. The idea behind exponential smoothing is to smooth the original time series data for forecasting future values.

The weighted exponential smoothing model (WES) is the most commonly used model for forecasting information from time series. This model is used when there are roughly horizontal information patterns, and there are no lengthy and temporary fluctuations. This means that the WES strategy is used to predict a time series when the time series data is at the normal stage.

(2) *K-Means Clustering Algorithm.* Clustering time series is one of the most difficult clustering problems in information mining time series. Subsequence time series is used by a sliding window to remove the subsequence of items, which is segment clustering from a single long time series. Another type of clustering is time-point clustering, which is used to cluster object time points based on a combination of temporal proximity and the similarity of their respective values. This sort of time series clustering is similar to the segmentation of time series. However, time-point clustering is distinct from segmentation, owing to the fact that in time-point clustering of all items, it is not appropriate to add to the cluster because some of the items are deemed noisy. In the clustering of subsequent time series, it is important to observe how the technique can be used to categorize a vast quantity of time series data on how they can generate significant results. A most recent study has focused on subsequent time series clustering to improve time series models. Our goal was to focus on clustering in the centroid to improve the model of the WES time series. It is important to note that to improve the WES time series model, our technique was more viable. The strategy of k -means clustering is one of the easiest unsupervised teaching methods to address the well-known issues of clustering. K -means clustering processes are very simple and easy to classify in a certain amount of clusters (suppose k clusters) in a given information set.

$$J = \sum_{i=1}^c \sum_{k=1}^{c_i} w_{ik} \|x^i - \mu_k\|^2, \quad (3)$$

where $\|x^i - \mu_k\|^2$ is the Euclidean distance, c_i is number of data points, and c is the number of clusters.

TABLE 1: Most significant 10 Google search terms.

Flu duration
Thermoscan
Fever flu
Influenza type a
Flu fever
Oscillococcinum
Braunthermoscan
Cold versus
Flu treating flu
Flu headache

3.3.2. Support Vector Machine Regression (SVMR). The support vector machine regression (SVMR) model is an increasingly common version of the support vector machine used for problems with regression. Although the Support Vector Machine algorithm is common in classification issues, SVMR is trained to generate numerical values for regression. The general formulation of SVM and SVMR algorithms is very different. The basic idea in both SVMR and SVM is to map data set X to a high-dimensional feature space F through a mapping function called kernel function π and to do linear regression in F [45, 46]. SVMR algorithm is essential to solve problems requiring many parameter estimates using traditional statistical methods. The SVM algorithm is used to classify data by using ε -insensitive. For the SVM algorithm, which uses regression purposes, the main idea was to find function $f(x)$ that has a deviation from the reality obtained target y_i for the training data. The main principle is the same as the SVM classification, but we have a new function that can be minimized. In the ε -insensitive support vector regression, the main goal is to find a function $f(x)$ that has a deviation from the actually obtained target y_i for all training data.

$$f(x) = wx + b, \quad w \in X, b \in R. \quad (4)$$

For this equation, we have to solve the following problem:

$$\min \frac{1}{2} \|w\|^2. \quad (5)$$

Subject to

$$\begin{aligned} y_i - wx_i - b &\leq \varepsilon, \\ wx_i + b - y_i &\leq \varepsilon. \end{aligned} \quad (6)$$

If the problems are not feasible, we need to introduce the slack variables ξ_i, ξ_i^* as it is called soft margin:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^I (\xi_i + \xi_i^*). \quad (7)$$

Subject to

$$\begin{aligned} y_i - wx_i - b &\leq \varepsilon + \xi_i, \\ wx_i + b - y_i &\leq \varepsilon + \xi_i^*, \quad \xi_i, \xi_i^* \geq 0, C > 0, \end{aligned} \quad (8)$$

For determination, the trade-off between the flatness of $f(x)$ by using C , the amount up for deviations is larger than

tolerance. This case is called ε -insensitive loss function $|\xi|\varepsilon$ and this can be as follows.

Figure 3 displays the hyperplane of the SVM algorithm when the hyperplane separates the data into classification and regression purposes. The SVM algorithm is used for classification data; it is a very powerful machine learning algorithm for classification, and the SVM algorithm has the ability to solve the regression problem, as shown in Figure 3. Figure 4 displays the process of the SVMR model to predict influenza outbreaks using Google search queries.

3.3.3. Artificial Neural Network Using Particle Swarm Optimisation (ANNPSO). Particle swarm optimisation was developed for a global optimisation system, PSO, which is a group based on a stochastic optimisation method for nonstop nonlinear capacities. In correlation with other metaheuristics, PSO has acquired prevalence and is indicated plainly to be successful, and it focused on enhancement calculation. Every part of the PSO technique has been known as particle flies around the multidimensional search space with a velocity, which is constantly raised to date by the particle's own particular experience and the experience of the particle's neighbours or the experience of the whole swarm. It implies two errors of the PSO algorithm are created: PSO with a neighbourhood in the global and PSO method with neighbourhood overall worldwide. As indicated by the global surroundings, every particle moves towards its best past position and towards the best particle in the whole swarm, called gbest demonstrate [47, 48]. Furthermore, as indicated by the local disparity called lbest, every particle moves towards its best past position and towards the best particle in its limited neighbourhood. While PSO has a memory of the past, the learning of a good solution is kept by all particles. Particles cooperate in a helpful way to share data in the swarm. The particle swarm optimisation (PSO) algorithm is based on a velocity update and position update. Velocity updates the following equation:

$$v_i(t+1) = wv_i(t) + c_1r_1(p_i(t) - x_i(t)) + c_2r_2(p_g(t) - x_i(t)). \quad (9)$$

The w is the inertia weight, c_1 and c_2 are the acceleration coefficients; uniformly distributed random numbers are r_1, r_2 and the values in the domain $[0, 1]$. This equation is updated by using the position update equation.

Position update:

$$x_i(t+1) = x_i(t) + v_i(t+1), \quad (10)$$

where random inertia weight was calculated according to the equation as follows:

$$w = 0.5 + \frac{\text{rand}()}{2.0}. \quad (11)$$

Artificial neural network (ANN) is a type of computational model that is regularly utilized in the fields of machine learning, software engineering, and other research disciplines. This computational model is composed to mirror the immense system of neurons in a brain. It is commonly

TABLE 2: Google search queries obtained from Google Correlate.

Flu duration	Symptoms flu
Thermoscan	Flu contagious period
Fever flu	Viral pneumonia
Influenza type a	Flu complications
Flu fever	Flu relapse
Oscillococcinum	Treat flu symptoms
Braunthermoscan	Flu in babies
Cold versus flu	Flu or cold
Treating flu	Flu quiz
Flu headache	What is tamiflu
Flu and fever	Duration of flu
Cold and flu symptoms	Flu germs
Tamiflu and breastfeeding	Human temperature
Flu in kids	Cold flu
Flu lasts	Tamiflu liquid
Tylenol flu	Treating flu symptoms
is tamiflu safe	Flu remedy
Flu symptoms last	Contagious flu
Anasbarbariae	Ear thermometer
Flu how long	Flu type a
Flu recovery	Cold or flu
Flu versus cold	Flu nausea
Side effects of tamiflu	Common flu symptoms
Type a flu	Common cold symptoms
Flu how long does it last	Influenza a
Low body temperature	Flu sweating
Treatment of flu	Oscillo
Flu dizziness	Length of flu
Flu recovery time	Flu in infants
Cold vs. flu	CDC Tamiflu
Flu length	Dangerous fever
Tamiflu side effects	What is influenza
Flu vomiting	Tamiflu suspension
Tamiflu in pregnancy	When to take Tamiflu
Flu without fever	Should I take Tamiflu
Flu stomach	Rapid flu test
Low body	Mild flu
Flu cough	Flu in adults
Tamiflu and alcohol	Fever temperature
Flu a and b	Tamiflu prophylaxis
Flu chest pain	Flu fever duration
Tamiflu dose	is it a cold or the flu
Flu contagious	Homeopathic flu
What is influenza a	Flu and rash
Treating the flu	Flu vs. cold
Flu how long contagious	Viral flu
Taking temperature	Tamiflu drug
Flu in toddlers	Fever reducers
Cough flu	Treat flu
Flu back pain	How to get over the flu

utilized for issues that are hard to be unequivocally customised in view of its capacity to gain from cases. The type of ANN utilized this exploration, which is completely associated with feedforward that organizes where each input is associated equitably with all the hidden neurons. For simplicity and preparation speed purposes, only a single hidden layer was utilized in the system. PSO is a global search and population-based algorithm used to train neural networks, identify neural network architectures, adjust network learning parameters, and optimize network weights. PSO

avoids trapping at a minimum local level because it is not based on information about gradients [47]. PSO function in ANN is to obtain the best set of weights (particle position) where several particles try to move to obtain the best solution. The search space dimension comprises cumulative weights and prejudices. By following the personal best solution of each particle and the best global amount of the entire swarm, the algorithm finishes the optimisation. A population-based algorithm's success or failure depends on its ability to trade efficiently between discovery and

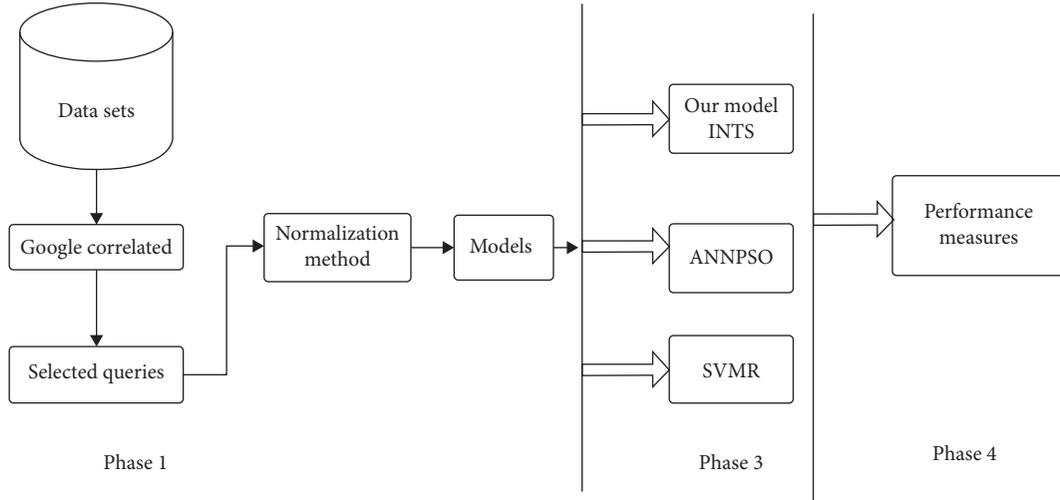


FIGURE 1: Generic framework of the proposed system.

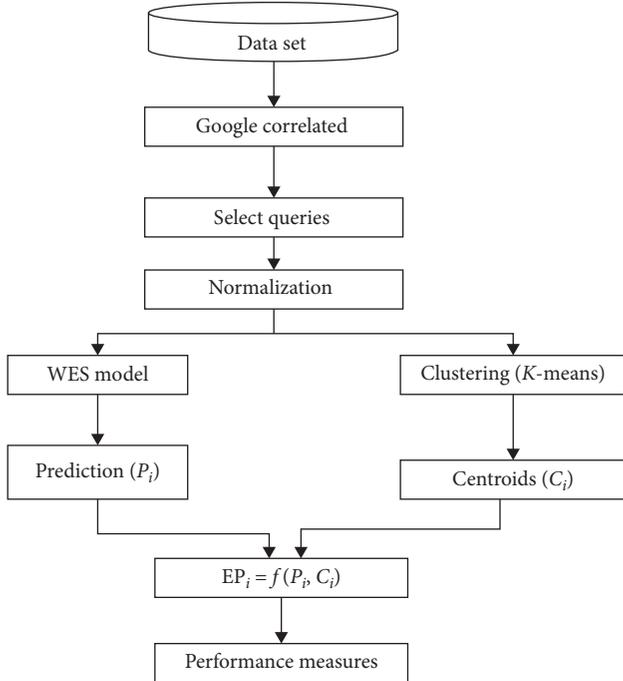


FIGURE 2: INTS model.

extraction. An inappropriate balance between exploration and extraction can result in a poor method of optimisation, which may suffer from premature convergence, local optimum trapping, and stagnation. Figure 5 shows the flow process of the ANNPSO model for predicting the influenza epidemic using Google search queries.

3.4. Performance Metrics. Four error indicators were used to evaluate the prediction model. The mean square error, root mean square error, and mean absolute error were used as performance indices. Those methods of standard indicators are defined as follows:

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^n (x_t - \bar{x}_t)^2, \quad (12)$$

where x_t is observed responses, \bar{x}_t are estimated responses, and N is the total number of observations.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{k=1}^n (x_t - \bar{x}_t)^2}, \quad (13)$$

where x_t is observed responses, \bar{x}_t are estimated responses, and N is the total number of observations.

$$\text{MAE} = |x_t - \bar{x}_t|, \quad (14)$$

where x_t and \bar{x}_t are the estimated and observed responses, respectively.

4. Results Analysis

Our analyses used the data from January 4, 2009 (week 1) to December 27, 2015 (week 52) across a total period of 312 weeks, covering 7 years of the CDC data. The CDC data are uploaded to Google Correlate, obtaining 100 search query terms that are related to the influenza epidemic. In total, 10 search terms with the highest correlation have been selected. The min-max method was used for normalisation purposes, and three experiments were conducted to obtain the prediction result. These three experiments are presented in the following section.

4.1. Results Analysis of the INTS Model. The Weighted Exponential Smoothing algorithm was applied to search terms obtained from Google correlate. The weighted exponential smoothing model depends on the α smoothing constant; it was then tested with values from 0.1 to 0.9. The MSE performance measure was scrutinized through the use of these parameters. The $\alpha=0.9$ parameter was selected as a

- (1) Cluster search queries obtained from Google correlate S_i using clustering techniques: k -means
- (2) Apply conventional time series prediction models, such as the WES model.
- (3) Obtain prediction P_i using WES model.
- (4) Modify P_i using C_i is centroid of the j^{th} cluster $EP_i = f(P_i, C_{-j})$

ALGORITHM 1: INTS algorithm.

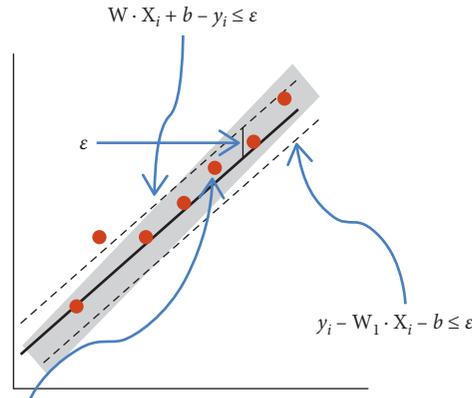


FIGURE 3: Hyperplane of SVM method.

smoothing constant. It was observed that $\alpha=0.9$ was appropriate for data prediction. Moreover, $\alpha=0.9$ is given fewer errors as compared to other parameters. To enhance the prediction of the conventional weighted exponential smoothing model, the k -means clustering algorithm is used. The first step was to determine the number of clusters for k -means clustering. The beginning was made up of eight clusters. After determining the existence of one cluster that had fewer objects, it was decided to reduce the cluster numbers until all clusters with more objects were obtained. Lastly, we determined that the five clusters were appropriate.

Then, it should be considered for centroids of cluster numbers. Each assigned object belongs to the specific cluster by centroids. The centroids were integrated with the results that have been achieved from the existing WES algorithm. The predictive capabilities of our intelligent model were compared with the existing GFT, ARGO, GFT + AR, AR (3), and naive models. Therefore, the comparison is presented by employing CDC real data. MSE, RMSE, and MAE were used to evaluate and estimate the performance of the INTS proposed model in comparison with the existing prediction models. The obtained results showed significant advantages for our proposed model. It was obvious that the INTS model is the most effective and robust predictor that can be used to enhance the prediction of the influenza epidemic using search terms. Table 3 summarizes the results obtained from our INTS model. The INTS model had shown the best performance in relation to MSE, INTS=0.0014, RMSE=0.0369, and MAE=0.0185. We noticed that the INTS model outperformed all existing models. We used correlated increments between the prediction and original CDC data. We observed that the INTS model had more

correlation when compared to other conventional models such as ARGO and GFT. The increment correlate obtained from the INTS model was 0.931. These findings indicate that the INTS model improved influenza epidemic prediction using search queries. Figure 6 illustrates the performance prediction of the intelligent time series model. Figure 7 shows the performance of the regression plot of the INT model.

4.2. Results Analysis of the ANNPSO Model. ANNPSO intelligent models have been implemented to predict the influenza epidemic using Google search terms. The Min-Max method is presented to scale the data to enhance the prediction models. Adapting the ANNPSO model was applied to develop a smart health care system. Since the weights of the ANN need to be optimized, the position of the particles in the PSO algorithm needs to be tracked. The issue space includes combinations of all weight values of the ANN algorithm. This search space consists of n -dimensions, where n is the total number of weights to optimize. Each particle has an n -dimensional location vector and speed vector. The particle swarm optimization is flying around this search space and creating the optimum weight set. The weights are allocated to the ANN while assessing the fitness of a particle in the PSO, and its predictive precision is discovered. This offers the particle's fitness. If fitness is the best so far for the particle, it will be taken as its personal best, and if it is the best so far for the swarm, it is considered the best global. The adapting model helps to improve prediction results. These particle swarms were used to improve the weight of the ANN approach. 25 particles were considered for 200 iterations.

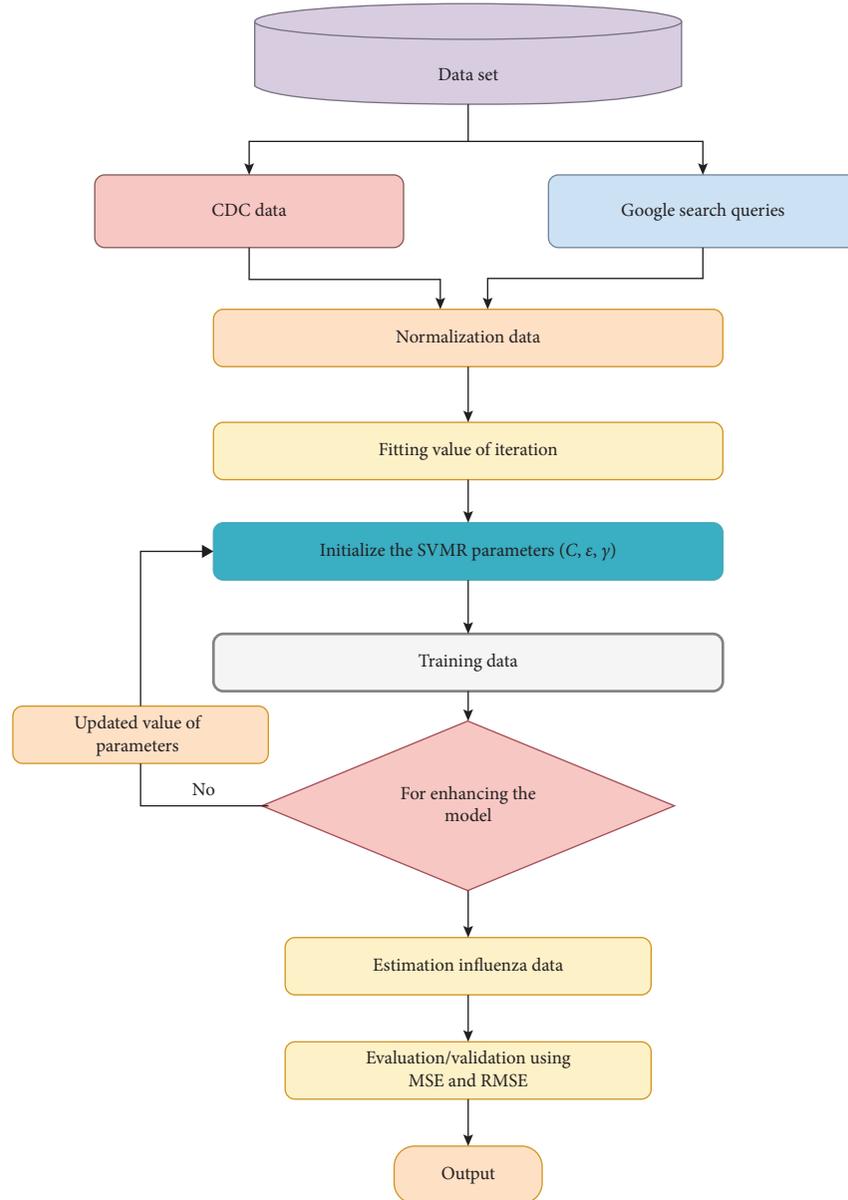


FIGURE 4: SVMR model is used to predict the influenza epidemic.

Table 3 summarizes the prediction results of ANNNPSO to predict influenza epidemics. It is noted that the adapting model obtained satisfying results. The prediction results were $MSE = 0.0024$ and $RMSE = 0.493$, $MAE = 0.25$, and $R = 0.94\%$. The obtained results have proved that Google search queries have the strongest relationship with clinical data. Figures 8 and 9 display the performance of the ANNNPSO model.

4.3. Results Analysis of the SVMR Model. Table 3 demonstrates the prediction results obtained from adapting the SVMR model. It is reported that the proposed model has performed good results. The support vector machine algorithm was applied to predict influenza, and we used the RBF kernel. The RBF kernel function has robust efficiency compared with other SVM functions. The kernel parameter

values C, γ, ϵ were tested to attain the best performance by changing parameter values. The optimum parameter values were selected according to the lowest obtained errors. The prediction results show that there is a relationship between clinical data and web search terms. According to the MSE, RMSE, MAE, and $R = 0.99$ obtained results of 0.136, 0.369, and 3.888, it is indicated that clinical data has more impact on the web search. Thus, Figures 10 and 11 exhibit the estimation performance of the SVMR model for predicting the influenza outbreak.

5. Discussion

In the present research paper, some significant implications have been presented for estimating and predicting the influenza epidemic at the local and national levels of USA

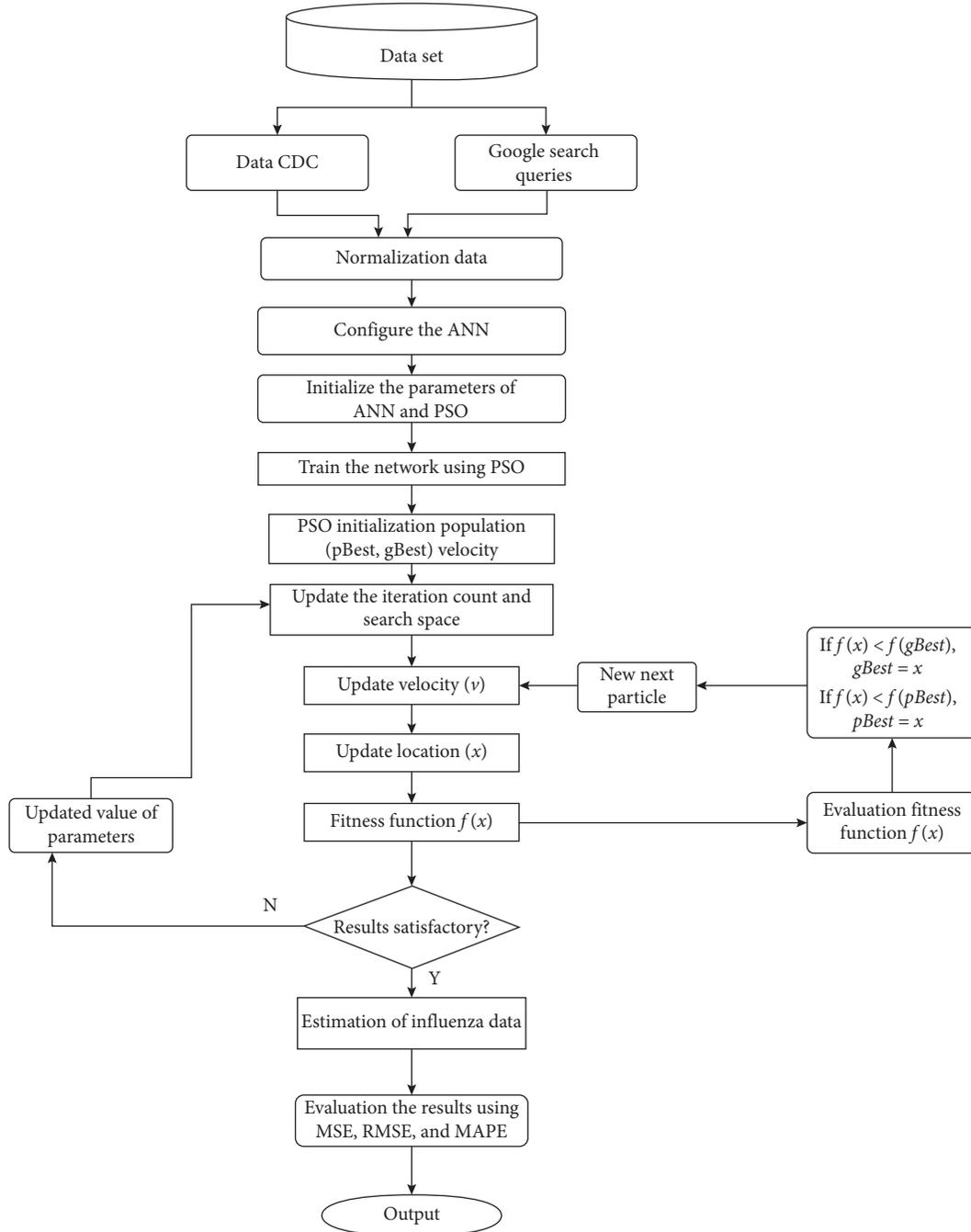


FIGURE 5: ANNPSO model used to predict the influenza epidemic.

TABLE 3: Results of prediction models for predicting influenza epidemics using Google search queries.

Models	MSE	RMSE	MAE	Correlation
INTS	0.0013	0.0369	0.0185	0.97
ANNPSO	0.0024	0.0493	0.259	0.94
SVMR	0.136	0.369	03.88	0.99

influenza data. In addition, early and precise detection of influenza outbreaks can assist in advising attempts to reduce the spread and effects of influenza outbreaks. The government can educate vaccination campaigns against influenza

outbreaks at the local and national levels. Having a precise scheme for monitoring influenza forecasts is especially useful for preventing and controlling the spread of influenza to other areas of the nation. Despite the fact that there is a

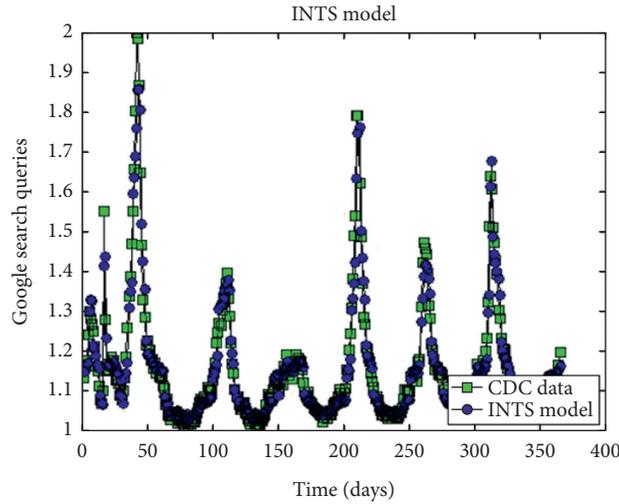


FIGURE 6: Performance of the INTS model.

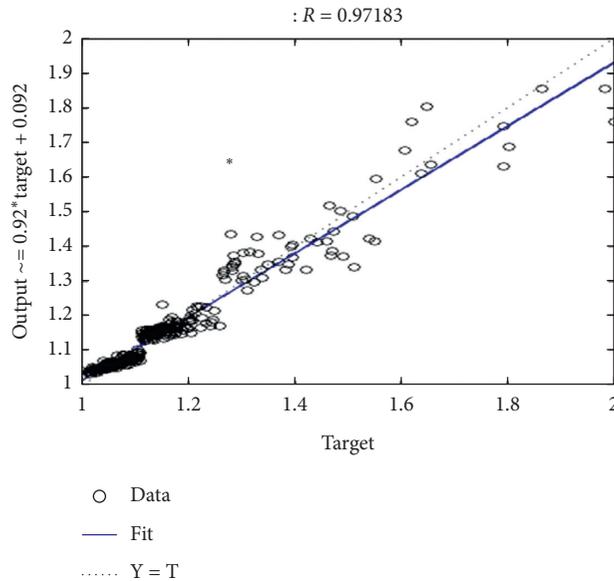


FIGURE 7: Performance regression plot of the INTS model.

trend towards modernizing surveillance of the influenza epidemic, the current standard models for influenza surveillance have documented shortcomings, including low sensitivity and less accuracy. Consequently, the need to take steps to improve influenza surveillance has been well acknowledged. Three adapting algorithms, namely, our INTS model, ANNPSO, and SVMR models, were implemented to predict the influenza epidemic. Hence, in the present study, we implemented an INTS model that can help to enhance the accuracy of influenza epidemic prediction. Our idea focuses on the centroids of clustering to improve the existing WES model and the time series model for predicting the influenza epidemic. The INTS model, assisted by a conventional time series prediction model and an appropriate machine intelligence approach, was used to predict influenza outbreaks using the Google search pattern. It is observed that

our model is more feasible for improving the time series model to predict the influenza epidemic. Figure 8 displays the correlation of INTS, and it is observed that the percentage of correlation is $R = 0.97$. The prediction results of this research demonstrate that, in general, models are built by Google search queries for estimating and predicting influenza outbreaks.

Furthermore, we have applied the adapting ANNPSO and SVMR models, which have performed better; it is noted that prediction errors are the lowest. These adapting models are compared with alternative models; they perform better for predicting influenza. These models were able to satisfactorily estimate true influenza outbreaks according to the official influenza case counts reported by the CDC for either a whole period or a seasoning period. The INTS model was compared with different existing models. The existing

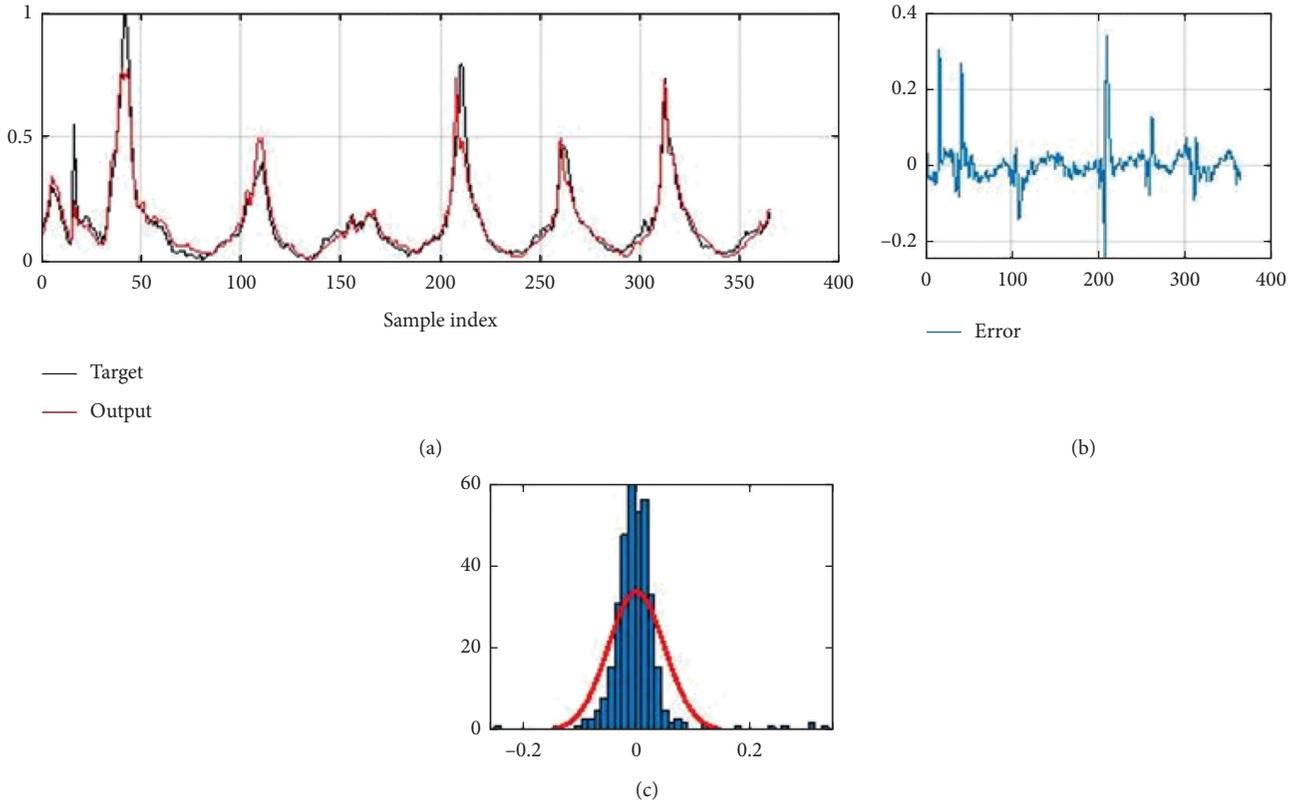


FIGURE 8: Performance of the ANNPSO model. (a) Train data. (b) MSE = 0.0024304, RMSE = 0.049299, and MAE = 0.2594. (c) Error St. D. = 0.049355.

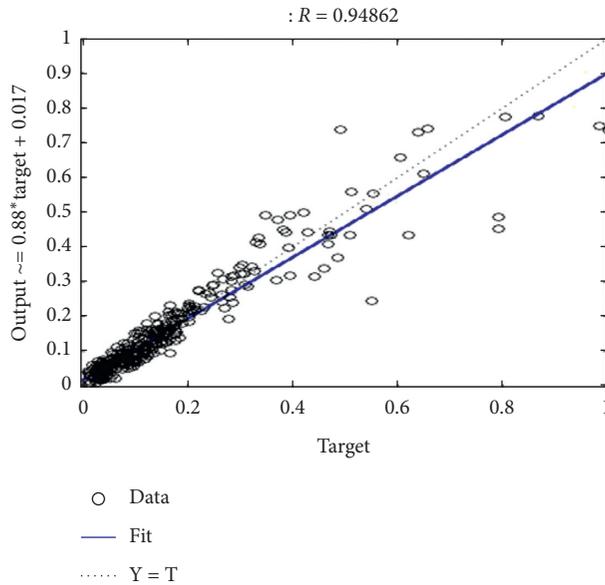


FIGURE 9: Performance regression plot of the ANNPSO model.

prediction models with different input data sets have used the pattern of the whole period and seasonal period. But, the INST model is used for one input data, with a whole data set. The results obtained from the intelligent model were compared with different existing models using different

input data sets throughout the period and the season. We observed that the INTS model outperformed all other alternative models of different input data with respect to MSE, RMSE, MAE, and increment correlate. Table 4 shows the perdition results of adapting the models against the existing

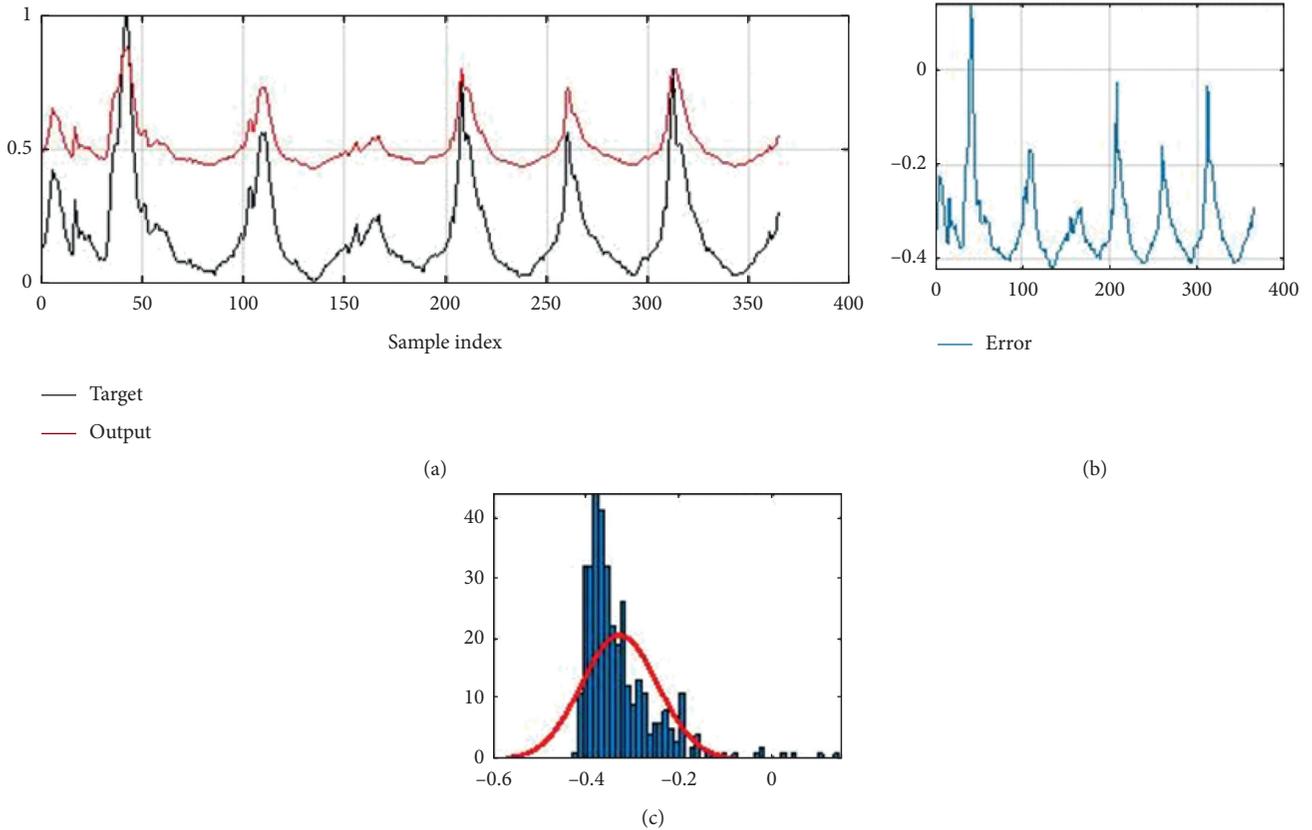


FIGURE 10: Performance of the SVMR model. (a) Train data. (b) MSE = 0.11444, RMSE = 0.33829, and MAE = 3.8887. (c) Error St. D. = 0.081002.

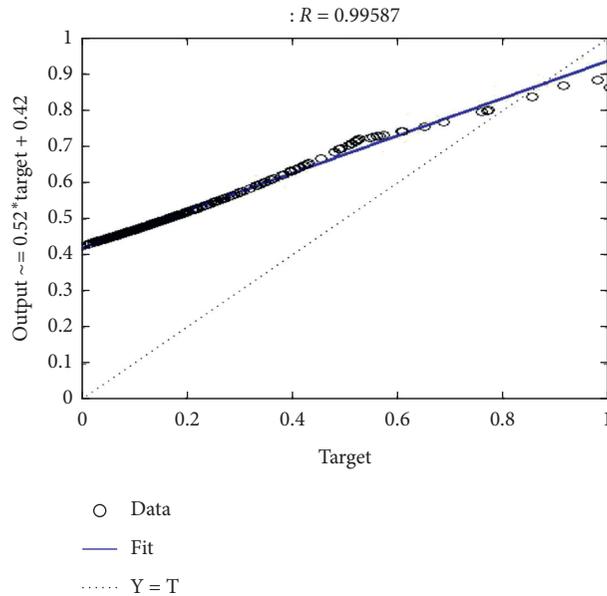


FIGURE 11: Performance regression plot of the SVMR model.

proposed system. The results of the INTS model are 0.0013, 0.0369, and 0.0185, according to MSE, RMS, and MAE. These results have been compared with other alternative strangest models, which are ARGO = 0.3696, GFT = 4.9106,

AR model = 0.915, naive model = 0.1211, and proposed adapting models ANNPS = 0.0024 and SVMR = 0.136 with respective to MSE metric. According to the RMSE evaluation metric, the INTS model = 0.0369, and other alternative

TABLE 4: Comparison of different models with the INTS model for the prediction of influenza epidemics using Google search queries.

Models	MSE	RMSE	MAE	Correlation of increment
INTS model (our model)	0.0013	0.0369	0.0185	0.9718
ARGO model [49]	0.3697	0.608	0.649	0.758
GFT	4.911	2.216	1.834	0.706
GFT + AR (3) [12]	0.8372	0.915	1.052	0.690
AR (3) model [9]	0.9157	0.957	0.925	0.385
Naive model [9]	0.1211	0.348	0.201	0.436
ANNPSO (proposed model)	0.0024	0.049	0.25	0.94
SVMR (proposed model)	0.136	0.369	3.88	0.99

models like ARG0 = 0.608, GFT = 2.216, AR model = 0.95, and proposed adapting models ANNPSO = 0.049 and SVMR = 0.369. Furthermore, INTS obtained fewer errors with regard to the MAE measure, INTS = 0.0231, and the conventional models, such as ARG0 = 0.649 and GFT = 1.834, whereas the ARG0 = 0.758, AR model = 0.925 and proposed adapting models ANNPSO = 0.25 and SVMR = 3.88. However, we have also noticed that the results provided by the INTS model have more efficiency in comparison to all alternative models, such as the GFT and ARG0 models and adapting models ANNPSO and SVMR with different periods. Further, the important advantage of predicting Google Internet search activities behaviour is the standing possibility for earlier detection. This is significant because taking notice by clinical or official health bodies is often delayed until there is an investigation of the prediction process. Furthermore, seeking significant health information on the web search activities before, or even instead of medical visits, can help detect earlier stages of the illness. It is therefore imperative to pay attention to the use of Google search queries for developing models that can help detect influenza in its earlier stages. This can be done by using CDC data that we have obtained from USA patients. Demonstrating and combining a robust, dynamic, and more accurate methodology to predict the influenza epidemic is our intention. We have concluded that the INTS model is stronger and more robust in comparison with other alternative models, such as ARG0 and GFT. The adapting model INTS is more efficient and robust in improving the healthcare system—the strength of the INTS model. The prediction results of the INTS model demonstrated the superiority of our model with respect to accuracy and strength compared to all alternative influenza prediction models over Google searches. The methodology of the proposed system can help improve the healthcare system by using Google search queries.

6. Conclusions

Web activities are an important source for obtaining health information. Consequently, web searches provide vital information regarding numerous infectious disease activities. For example, an influenza epidemic can occur when an infectious disease quickly spreads to many people. As a further instance, the web can efficiently perform a thorough examination of the relationship between search queries of influenza and actual influenza occurrence. Three adapting

prediction models, namely INTS, ANNPSO, and SVMR, were presented to improve the prediction of the influenza epidemic by using Google search queries. The methodology of these proposed models outperformed other existing models and provided higher accuracy and robustness in predicting influenza. The former models were originally implemented to predict influenza using Google search terms. However, the novelty of the proposed research is the development of the INTS model, which is a new model for predicting the influenza epidemic. The prediction results demonstrate that the proposed INTS model can be effectively employed to predict influenza outbreaks using Google search queries. A comparative prediction result between GFT, ARG0 models, and the present SVMR and ANNPSO models is presented. It has been observed that the results of the alternative model are 0.0013, 0.0369, 0.0185, and 0.97, in accordance with MSE, RMSE, MAE, and correlate of increment performance measures. Respectively, it is also observed that the INTS model is more satisfying in comparison to the existing models like ARG0 and GFT.

Data Availability

The researchers can collect the data from <https://gis.cdc.gov/grasp/fluview/FluView8.html> and <https://www.google.com/trends>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the Deanship of Scientific Research at King Faisal University for the financial support under Nasher Track (Grant no. 186344).

References

- [1] “Overview of influenza surveillance in the United States, centers for disease control and prevention,” 2019, <https://www.cdc.gov/flu/weekly/overview.htm>.
- [2] F. Wang, H. Wang, K. Xu et al., “Regional level influenza study with geo-tagged twitter data,” *Journal of Medical Systems*, vol. 40, no. 8, p. 189, 2016.
- [3] Y. Wang, K. Xu, Y. Kang, H. Wang, F. Wang, and A. Avram, “Regional influenza prediction with sampling twitter data and PDE model,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 678, 2020.

- [4] D. A. Broniatowski, M. J. Paul, and M. Dredze, “National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic,” *PLoS One*, vol. 8, Article ID e83672, 2013.
- [5] M. Smith, D. A. Broniatowski, M. J. Paul, and M. Dredze, “Towards real-time measurement of public epidemic awareness: monitoring influenza awareness through twitter,” in *AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*, George Washington University, Washington, DC, USA, 2016.
- [6] K. Wilson and J. S. Brownstein, “Early detection of disease outbreaks using the Internet,” *Canadian Medical Association Journal*, vol. 180, no. 8, pp. 829–831, 2009.
- [7] M. W. Davidson, D. A. Haim, and J. M. Radin, “Using networks to combine “big data” and traditional surveillance to improve influenza predictions,” *Scientific Reports*, vol. 5, no. 1, 2015.
- [8] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, “Detecting influenza epidemics using search engine query data,” *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.
- [9] Q. Yuan, E. O. Nsoesie, B. Lv, G. Peng, R. Chunara, and J. S. Brownstein, “Monitoring influenza epidemics in China with search query from baidu,” *PLoS One*, vol. 8, no. 5, Article ID e64323, 2013.
- [10] J. Shaman and A. Karspeck, “Forecasting seasonal outbreaks of influenza,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 50, pp. 20425–20430, 2012.
- [11] E. H. Chan, V. Sahai, C. Conrad, and J. S. Brownstein, “Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance,” *PLoS Neglected Tropical Diseases*, vol. 5, no. 5, p. e1206, 2011.
- [12] M. Santillana, D. W. Zhang, B. M. Althouse, and J. W. Ayers, “What can digital disease detection learn from (an external revision to) Google flu trends?” *American Journal of Preventive Medicine*, vol. 47, no. 3, pp. 341–347, 2014.
- [13] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, “Multiscale mobility networks and the spatial spreading of infectious diseases,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21484–21489, 2009.
- [14] V. Colizza, A. Barrat, M. Barthelemy, A. Valleron, and A. Vespignani, “Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions,” *PLoS Medicine*, vol. 4, no. 1, p. e13, 2019.
- [15] D. Balcan, H. Hu, B. Goncalves et al., “Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility,” *BMC Medicine*, vol. 7, no. 1, 2009.
- [16] S. Eubank, H. Guclu, V. S. Anil Kumar et al., “Modelling disease outbreaks in realistic urban social networks,” *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.
- [17] N. M. Ferguson, D. A. T. Cummings, C. Fraser, J. C. Cajka, P. C. Cooley, and D. S. Burke, “Strategies for mitigating an influenza pandemic,” *Nature*, vol. 442, no. 7101, pp. 448–452, 2006.
- [18] D. M. Goedecke, G. V. Bobashev, and F. Yu, “A stochastic equation-based model of the value of international air-travel restrictions for controlling pandemic flu,” in *Proceedings of the 2007 Winter Simulation Conference*, Washington, DC, USA, 2007.
- [19] M. L. C. D. Atti, S. Merler, C. Rizzo et al., “Mitigation measures for pandemic influenza in Italy: an individual based model considering different scenarios,” *PLoS One*, vol. 3, no. 3, p. e1790, 2008.
- [20] Y. Zhang, G. Milinovich, Z. Xu et al., “Monitoring pertussis infections using internet search queries,” *Scientific Reports*, vol. 7, no. 1, 2017.
- [21] F. Rohart, G. J. Milinovich, S. M. R. Avril, K.-A. L. Cao, S. Tong, and W. Hu, “Disease surveillance based on Internet-based linear models: an Australian case study of previously unmodeled infection diseases,” *Scientific Reports*, vol. 6, no. 1, 2016.
- [22] V. Lampos, A. C. Miller, S. Crossan, and C. Stefansen, “Advances in nowcasting influenza-like illness rates using search query logs,” *Scientific Reports*, vol. 5, no. 1, 2015.
- [23] S. Cho, C. H. Sohn, M. W. Jo et al., “Correlation between national influenza surveillance data and Google trends in South Korea,” *PLoS One*, vol. 8, no. 12, 2013.
- [24] A. F. Dugas, M. Jalalpour, Y. Gel et al., “Influenza forecasting with Google flu trends,” *Online Journal of Public Health Informatics*, vol. 5, no. 1, 2013.
- [25] S. Towers, S. Afzal, G. Bernal et al., “Mass media and the contagion of fear: the case of ebola in America,” *PLoS One*, vol. 10, no. 6, 2015.
- [26] D.-C. Huang and J.-F. Wang, “Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors,” *Science of The Total Environment*, vol. 612, pp. 1293–1299, 2018.
- [27] H. Tenkanen, E. D. Minin, V. Heikinheimo et al., “Assessing the usability of social media data for visitor monitoring in protected areas,” *Scientific Reports*, vol. 7, no. 1, 2017.
- [28] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, “Forecasting the onset and course of mental illness with twitter data,” *Scientific Reports*, vol. 7, no. 1, 2017.
- [29] T. H. Aldhyani, A. S. Alshebami, and M. Y. Alzahrani, “Soft clustering for enhancing the diagnosis of chronic diseases over machine learning algorithms,” *Healthcare Engineering*, vol. 2020, no. 4984967, pp. 16–16, 2020.
- [30] M. Santillana, E. O. Nsoesie, S. R. Mekaru, D. Scales, and J. S. Brownstein, “Using clinicians’ search query data to monitor influenza epidemics,” *Clinical Infectious Diseases*, vol. 59, no. 10, pp. 1446–1450, 2014.
- [31] 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, Article ID e55205, 2013.
- [32] G. J. Milinovich, S. M. R. Avril, A. C. A. Clements, J. S. Brownstein, S. Tong, and W. Hu, “Clinical infectious diseases,” *Healthcare Epidemiology*, vol. 47, no. 11, pp. 1–6, 2008.
- [33] G. J. Milinovich, S. M. R. Avril, A. C. A. Clements, J. S. Brownstein, S. Tong, and W. Hu, “Using internet search queries for infectious disease surveillance: screening diseases for suitability,” *BMC Infectious Diseases*, vol. 14, no. 1, 2014.
- [34] S. Cook, C. Conrad, A. L. Fowlkes, and M. H. Mohebbi, “Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic,” *PLoS One*, vol. 6, no. 8, Article ID e23610, 2011.
- [35] E. O. Nsoesie, J. S. Brownstein, N. Ramakrishnan, and M. V. Marathe, “A systematic review of studies on forecasting the dynamics of influenza outbreaks,” *Influenza and Other Respiratory Viruses*, vol. 8, no. 3, pp. 309–316, 2013.
- [36] J.-P. Chretien, D. George, J. Shaman, R. A. Chitale, and F. E. McKenzie, “Influenza forecasting in human populations: a scoping review,” *PLoS One*, vol. 9, no. 4, Article ID e94130, 2014.
- [37] B. Alkouz, Z. A. Aghbari, and J. H. Abawajy, “Tweetlenza: predicting flu trends from twitter data,” *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 273–287, 2019.

- [38] <https://www.google.com/trends/correlate>.
- [39] T. H. Aldhyani, A. S. Alshebami, and M. Y. Alzahrani, "Soft computing model to predict chronic diseases," *Journal of Information Science and Engineering*, vol. 36, no. 2, pp. 365–376, 2020.
- [40] F. W. Alsaade and M. H. Al-Adhaileh, "Developing a recognition system for classifying covid-19 using a convolutional neural network algorithm," *Computers, Materials & Continua*, vol. 680, no. 1, pp. 805–819, 2021.
- [41] M. Alrasheed, M. H. Al-Adaileh, A. A. Alqarni, and M. Y. Alzahrani, "Deep learning and holt-trend algorithms for predicting covid-19 pandemic," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2141–2160, 2021.
- [42] M. A. Aleid, M. Rahmouni, N. Alsharif et al., "Modelling the psychological impact of covid-19 in Saudi Arabia using machine learning," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2029–2047, 2021.
- [43] <https://gis.cdc.gov/grasp/fluview/fluportaldashboard.htm>.
- [44] <https://www.cdc.gov/flu/weekly/>.
- [45] H. Yang, L. Chan, and I. King, "Support vector machine regression for volatile stock market prediction," *Intelligent Data Engineering and Automated Learning—IDEAL 2002 Lecture Notes in Computer Science*, pp. 391–396, Springer, Berlin, Germany, 2002.
- [46] Y.-C. Guo, D.-X. Niu, and Y.-X. Chen, "Support vector machine model in electricity load forecasting," in *Proceedings of the 2006 International Conference on Machine Learning and Cybernetics*, Jeju Island, South Korea, 2006.
- [47] S. Yogi, K. R. Subhashini, and J. K. Satapathy, "A PSO based functional link artificial neural network training algorithm for equalization of digital communication channels," in *Proceedings of the 2010 International Conference on Industrial and Information Systems (ICIIS)*, pp. 107–112, Mangalore, India, 2010.
- [48] Y. Zhang and L. Wu, "Crop classification by forward neural network with adaptive chaotic particle swarm optimization," *Sensors*, vol. 11, no. 5, pp. 4721–4743, 2011.
- [49] S. Yang, M. Santillana, and S. C. Kou, "Accurate estimation of influenza epidemics using Google search data via ARGO," *Proceedings of the National Academy of Sciences*, vol. 112, no. 47, pp. 14473–14478, 2015.