JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 365-376 (2020)
DOI: 10.6688/JISE.202003_36(2).0014

Soft Computing Model to Predict Chronic Diseases

THEYAZN H. H. ALDHYANI[1], ALI SALEH ALI ALSHEBAMI[1]
AND MOHAMMED YAHYA ALZAHRANI[2]
[1]*Community College of Abqaiq*
King Faisal University
Ahsaa, 31982 Kingdom Saudi Arabia
E-mail: {taldhyani; aalshebami}@kfu.edu.sa
[2]*Information Technology Department*
Albaha University
Albaha, 65527 Kingdom Saudi Arabia
E-mail: msawileh@bu.edu.sa

The World Health Organization (WHO) has reported that non-communicable diseases are too risky as one of the serious diseases that threaten this world. The chronic diseases are extremely complex, so collaboration with ecological, biological and behavioral has given challenges for researchers and developers to predict non-communicable diseases. Digital surveillance system aimed to detect search queries that help to improve the awareness and timeline of predicted health outbreak. The purpose of the research study is to use Google Trend data for predicting chronic diseases. In this study, soft computing algorithm is applied to map the web search activity behavior of the population to prevent chronic disease risk factors. The Google Trend search activity is used to identify relevant web search activity, so the study period was taken from the first week of January 2017 to last week of December 2017 as it consisted of four chronic diseases namely asthma, heart, diabetics, and kidney. The clinical data of chronic diseases were collected from the Centers for Disease Control and Prevention (CDC) to test and evaluate the proposed model. The Adaptive Neuro-Fuzzy Inference System (ANFIS) algorithm was used to map web search activity from Google Trend using CDC clinical data. The consideration data in 2017 has shown there is strongly correlation between web search activity and clinical data. The standard evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were applied to evaluate the performance of the proposed model. The experimental results have shown that there is a relationship between internet search queries and clinical data, and thus prediction errors are very less. The high predictive validity of web search queries for chronic diseases have given the possibility to consider the population web information in order to predict Non-Communicable Disease (NCD) risk for avoiding and spreading in a large area scale. It is concluded that the proposed system can help to detect and predict the non-communicable diseases in the earliest stage.

Keywords: chronic diseases, soft computing, Google trend data, intelligent model, machine intelligent

1. INTRODUCTION

The rapid adoption of the internet has opened a gate for developing and enhancing healthcare. Numbers of researchers have used a huge of data that comes from the internet and social media such as Twitter or Facebook for discovering a novel method to diagnose the diseases. The language patterns from the internet and social media are demon-

Received September 30, 2019; revised October 2&11, 2019; accepted October 14, 2019.
Communicated by Osamah Ibrahim Khalaf.

strated to analyze and predict the chronic disease, finding out the behavioral habits are increased the diseases. Understanding population, behavior and trends of chronic disease risks are identified by using Web search activity data. The chronic disease risks have been detected by using search activity data to examine the data that has been submitted to the health official, and thus these search activity data have the same trend with the examination of data. The researchers have compared web search data relating to the main key modifiable risk factors of chronic disease with clinical population data from the US Centre for Disease Control. Developing real-time surveillance can provide a proxy for clinical population data and real-time web search data to enhance the healthcare system.

The chronic diseases such as cancer, asthma, heart, and diabetics are Non-Communicable Diseases (NCD) as compared with another global disease that is an extremely serious type of global disease. The chronic diseases spend very slowly and thus they are not given any signal that body of humans suffering from these diseases, therefore, the World Health Organization (WHO) has reported the chronic disease is one of the highest grave diseases that threaten human life in this world. The WHO provides the framework for sympathetic risks of chronic diseases such as a raised blood sugar, raised blood pressure dyslipidemia, overweight, obesity, and abnormal lung function. Besides, they illuminate the behavioral habits from environmental factors that belong to increasing chronic diseases such as factors (unhealthy diet, physical inactivity, tobacco and alcohol use, air pollution, age, and heredity). In addition, there are the same environmental factors that can grow the chronic diseases like (globalization, urbanization, population aging, and social determinants). The difficulty of chronic disease has given a challenge to most of the researchers for developing and designing an appropriate system to prevent spending on a large scale. Designing a surveillance system by using real-time data is to give more attention to this study. However, there is still some difficulty in collecting data due to the limited usefulness of the time that lags the current inherent data gathering approaches for measuring population risk. Therewith, the researchers have used web search data to convert into real-time data that can help to prevent the spent chronic diseases in a large area. The web search can provide insight into the problems of the population by analyzing search patterns from web activities. Google search volumes are used as a proxy to analyze the behavior of population such as fitness, diet, weight loss, and smoking for detection chronic disease [18, 19]. Similarly, considering the web search activity data as useful information, it becomes a proxy for chronic diseases are related to risk behavior and habits. Further, these risk behavior and habits provide a step to avoid risks and expend for chronic diseases [20]. In this present search, work uses a machine- learning algorithm to predict web search activity and behavior as a proxy for chronic disease risk factors. The novelty of the research is used web search activity patterns to detect and predict non-communicable diseases. Meanwhile, a soft computing algorithm is proposed to improve the prediction process. The paper is organized as follows in Section 1 presented an introduction. The significant of study is discussed in Section 2. In Section 3 presents background of study material and methods is analyzed in Section 4. Finally, the paper is stopped up with conclusion in Section 5.

2. BACKGROUND OF STUDY

A. Jurgen [1] proposed the Auto regression model to predict the influenza epidemic.

Internet search activity was used to enhance short-term forecasts. Some researchers demonstrated search queries data to estimate and predict the influenza epidemic. S.-Y. Shin *et al.* [2] Used the correlation method to find out the relation between National Influenza Surveillance Data and search terms obtained from smartphones and computers in South Korea [2]. For using Google trend search terms, A. Valdivia *et al.* [3] used a correlation method to predict influenza by using Europe data. Moreover, J. Ginsberg *et al.* [4] proposed a linear regression model to find a correlation between search queries and CDC data from USA. In addition, S. Cho *et al.* [5] Presented correlation method to predict influenza epidemic by using south Korea clinical data with search queries obtained from Google trend for south Korea. D. W. Seo *et al.* [6] proposed a cumulative query method to predict influenza diseases of Korea. The authors correlated Korea center health data with queries terms obtained from a search engine of Korea. Using Twitter data for predicting healthcare diseases, there are numbers of research have introduced like D. A. Broniatowski *et al.* [7] used correlation method for discovering the relation between tweets and clinical data obtained from CDC, R. Nagar *et al.* [8] proposed linear regression model to predict influenza using twitter data with using CDC data, J. C. Santos *et al.* [9] authors proposed classification algorithm namely Navies Bays and SVM to discriminate the tweets that submitted by illness peoples. Further, the linear regression model is employed to find the relation between the tweets' and clinical data from CD. Furthermore, the researchers have used Wikipedia data to predict the diseases for improving the health system. N. Generous *et al.* [10] used correlation method to predict global disease from WHO, the Wikipedia data are used to find the relation between the data obtained from search web activity and health data from WHO. D. J. McIver *et al.* [11]. The Poisson model is presented to predict influenza outbreak, authors used Wikipedia search data with CDC data. P. M. Markey *et al.* [12] used a linear model to predict chronic diseases. The Google trend data is considered to correlate with CDC clinical data. Authors have taken behaviors of chronic diseases like smoking and obesity. C. Li *et al.* [13] presented a correlation method to find the relation between the search quires submitted by illness peoples with CDC. M. Santillana *et al.* [14] demonstrates an alternative methodology to enhance the GFT model with guarantee value for digital disease detection broadly. The alternative model automatically selects specific queries to monitor updates of the model for an estimate influenza epidemic. M. Kang *et al.* [15], Q. Yuan *et al.* [16], M. Philip *et al.* [17] develop a model to predict influenza activity with the use of Internet search queries from influenza surveillance in China. Soft computing algorithms such as a neural network and fuzzy system have concerned to predict and analyzes the healthcare system.

3. MATERIAL AND METHODS

The current research work develops a predictive model for analyzing chronic disease risk based on web search activity and clinical data. The proposed model is used to identify the relation between web search quarries submitted by population and clinical data available in official health. For measuring the chronic diseases, the clinical data has collected from centers Disease Control and Prevention (CDC). These clinical data were obtained from a non-communicable disease surveillance system in the USA. In addition, these data sets were reported weekly percentage of patients who are suffering from chronic diseases. The Log transformation method is proposed can be normalized the data. An evaluation phase used the same metrics namely MSE, RMSE, and MAE for testing the proposed model. The detailed description of the proposed model will be discussed in the following subsections.

3.1 Data Sets

(A) Clinical Data

The clinical data is known as a dependent variable data set. The data set was collected from the Centers for Disease Control and Prevention (CDC) to operate the Behavioral Risk Factor Surveillance System (BRFSS). In this paper, the use of chronic disease benchmark for population risk prevalence is in one year. The training data has been obtained from the first week of January 2017 to the last week of December 2017. We have used four types of chronic diseases named Asthma, Diabetes, and heart diseases.

(B) Web Search Data

Web search data is defined as an independent variable. The web search activity data is search queries obtained from Google Trend. We have obtained web search queries that have been have submitted or searched by the population of the USA during the period of 2017. Such search queries related to chronic diseases are considered more important. These search queries have useful information for helping to detect or identify chronic diseases. It is emphasized only the search has named some queries of the related disease. The Google Trend supports the researchers to discover the highest search queries related to specific diseases. For obtaining these queries, we have just an entry name of chronic diseases in Google Trend. These search queries are considered as target queries for finding the relation between the search queries and patient registration in official health. The considered search terms related to chronic diseases have been investigated as most of the search terms have some similar names of chronic diseases. It is observed that 24 search terms are driven from 157 search terms; these 24 terms have similar names of chronic diseases. Finally, it is noted that these queries are most likely to be associated with chronic disease types by using the accumulated posts during the critical 1 year period (between the first weeks of 2017 to last week 2017). Table 1 shows search queries obtained from Google Trend.

3.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid of Multilayer Feed-Forward Neural Network and fuzzy logic. ANFI algorithm is constructed input-output mapping based on the initial given fuzzy logic system by using a neural network. It is used to predict and analyze the data with superior non-linear mapping that is used to map input space into output space. Furthermore, mapping two Fuzzy Inference Systems (FIS) are used in numbers of real-time applications. These are the Sugeno inference system and the Mamdani inference system, which are presented in the literature. The fuzzy rules for these two inference approaches are different in aggregation and defuzzification. The proposed model is used to predict chronic diseases by using web search, thus the steps of the soft computing model are shown in Fig. 1. The clinical data and web search are nor-

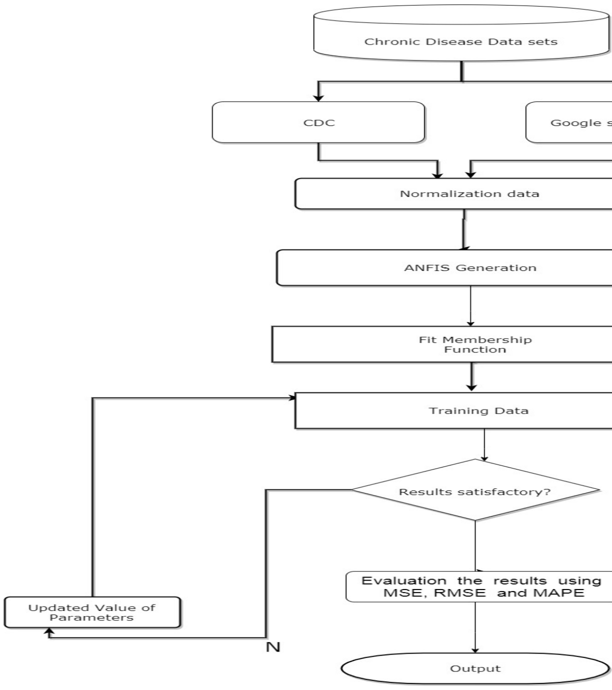


Fig. 1. Proposed model.

Table 1. Search queries of chronic diseases obtain from Google trend.

Search queries of asthma diseases	Search queries of heart diseases	Search queries of diabetics diseases	Search queries of kidney diseases
asthma	heart	diabetes type 2	kidney stones
allergy	heart disease	symptoms	kidney pain
allergy asthma	cardiovascular	diabetes symptoms	kidney disease
allergy and asthma	heart problems	type 1 diabetes	kidney stone
asthma attack	cardiovascular disease	what is diabetes	kidney infection
asthma symptoms	symptoms	gestational diabetes	kidney symptoms
asthma inhaler	cardiovascular system	sugar diabetes	symptoms
what is asthma	cvd	symptoms of diabetes	kidney failure
inhaler	heart disease symptoms	diabetes diet	kidney transplant
asma	what is heart disease	diabetes signs	kidney cancer
asthma cough	coronary heart disease	diabetic	kidney function
asthma treatment	congenital	signs of diabetes	back pain
icd 10 asthma	what is cardiovascular	diabetes test	liver
bronchitis	congenital heart disease	diabetes mellitus	kidney infection symptoms
symptoms of asthma	diabetes	diabetes causes	kidneys
asthmatic	cardiovascular risk	diabetes insipidus	kidney beans
allergies	heart attack	diabetes icd 10	kidney stones symptoms
Copd	symptoms of heart disease	diabetes treatment	kidney stone pain
asthma inhalers	blood pressure	diabetes association	chronic kidney disease
asthma test	heart disease causes	diabetes care	kidney stones pain
asthma causes	heart diseases	diabetes 2 symptoms	kidney problems
asthma exacerbation	cardiovascular health	pregnancy diabetes	uti
asthma medicine	heart failure	pre diabetes	kidney diet
allergy and asthma center	cardiovascular definition	diabetes definition	kidney stone symptoms
asthma medication	women heart disease	diabetes type 2 symptoms	kidney pain symptoms

malized by using the log function method then processed by the ANFIS model. The ANFIS model is employed the Gaussian input parameter membership function to analyze the patterns. The parameters of the ANFIS model is shown in Table 2. The relationship between clinical data and web search are used to generalize the relationship between the dependent variable (clinical data) and the independent variable (web search), for modeling the chronic disease data.

Table 2. Parameters of ANFIS.

Parameters for ANFIS by FCM algorithm	
Parameter cluster	2
Partition	2
Maximum number of iteration	200
Minimum imp	1e-5
Maximum number of epochs	200
Error Goat	0
Initial steps size	0.01
Step size decrease rate	0.9
Steps size increase	1.1

3.2.1 Modeling of ANFIS model

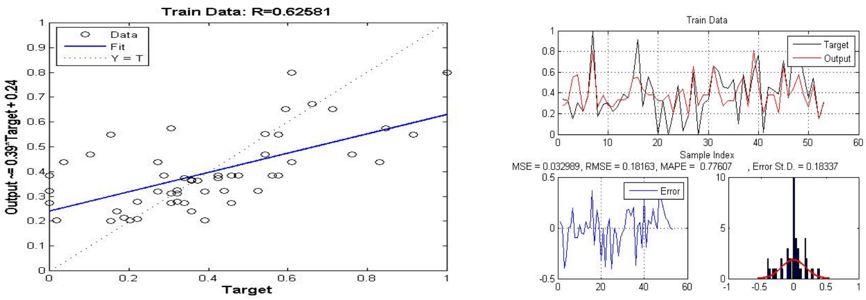
The ANFIS model is employed automatically to generate Gaussian shaped membership functions, the Fuzzy C-Means (FCM) algorithm is considered to analyze and predict chronic disease. This algorithm is generated fuzzy if-then rules for predicting training data. Fuzzy c-means clustering uses to cluster the data in which the object belongs to all the clusters with different membership values. The FCM method is proposed by Bezdek, the data is clustered into 12 clusters. The clustering technique in each data point belongs to a cluster and some degree that is called membership. The cluster estimates are obtained can be used to initialize iterative optimization based on clustering methods and model identification methods like ANFIS. In this research work, 15 clusters have been measured for the given one-year training data. The number of clusters should be equal to the numbers of fuzzy rules; this will represent the characteristic of cluster data. The parameters of FCM clustering with fuzzy rules are given in Table 3. At each epoch, an error measure is the sum of the squared difference between actual and desired output is reduced. When the values of the premise parameters are learned, the overall WQI is obtained as a linear combination of these parameters. This research work presents evidence regarding the web search quarries form Google trend that has the strongest relation with clinical data by using soft computing ANFIS model. The data collected from CDC. The period of the analysis was the first week of January 2017 to the last week of December 2017, taking time 52 weeks. Four chronic diseases namely asthma, heart, diabetics and Kidney are considered for finding the relationship with web data. We have searched for the similar term of each chronic disease with the help of Google Trend in the same period of clinical data. Google trend has produced web search terms for 52 terms in the USA and for each disease in the period of one year in 2017. The obtained search terms are presented in Table 2. The search terms were normalized by using log function, then by using fuzzy rules based on finding relation with CDC patients.

4. RESULTS AND ANALYSIS

The proposed research work demonstrates and combines web search queries with clinical data for predicting the chronic disease; it is observed that this methodology is a more robust dynamic for predicting chronic diseases based on an internet search. The non-linear soft computing model is presented to predict chronic diseases. The non-linear model is more robust as compared to the linear model for predicting healthcare data. The main target of the proposed model is to approve the there is any relation between the clinic data obtaining from the official health and search queries that have been submitted by the USA population. The research data sets have been collected from the CDC and Google Trend. This research work is implemented by using Matlab programming, four evaluation indicators are applied to test and evaluate the soft computing model. From the experimental analysis, it has found there accidental to predict non-communicable diseases by using web activity search; making the proposed model for the visible Log normalization method is used. The log method is used to scale the data. The research work can help to prevent spreading chronic diseases in large area by scaling and identifying the search term that has submitted by the population in the USA. As we have known the non-communicable diseases are one of danger disease and increasing by using wrong behaviors or habitats. By using web search, we can find all wrong behavior and season which increase the chronic diseases and find a specific solution for controlling and predicting the non-communication diseases. The presents research can help government to identify the when the specific chronic is speeded, can find some vaccination or nay solution for discontinuing spread of non-communicable epidemics. The proposed model is appeared to accurately predict the rank order of states across the USA for each of the NCD risk factors examined. The possible utility of this proposed model is twofold: first, the internet-based activity terms data are publicly available in approximate real time and considered for selecting the most significant internet terms that have related to non-communicable epidemics. The soft computing proposed model has provided instant feed- back to interference implementation and evaluation. Second fold is gathering the official clinical data from the CDC to analyze and discover the patterns for predicting. The proposed model can be generated ahead of measured and predicted data of numbers of the population survey. The web search collated in 2017 has proved that the information submitted by the population survey is most useful for improving the healthcare system in the USA. Table 3 shows the prediction results of the soft computing model for predicting non-communication diseases. It is observed that their strongest relation between clinic data from official health and web search terms form Google trend that has been searched and submitted by USA population. The predicted results of the asthma disease are shown that there is a correlation between the clinical data and web search activities. The obtaining results are 0.0331, 0.1819, 0.7874 and 0.1836 with respect to MSE, RMSE, MAE and ErroSTD. Fig. 2 (b) displays the time series plot of the proposed model for predicting asthma diseases, it is noted the predicted results similar to actual data. Moreover, the predicted errors are less and thus this approved that the search terms are clinical data in the same years have relation. Fig. 2 (a) shows the regression plot of proposed for predicting the asthma diseases, it is observed that most of the data point in the regression line, the correlation results $R = 0.625$. Furthermore, the prediction results of heart are to $MSE = 0.0180$, $RMSE = 0.1341$, $MAE = 0.0671$ and $ErroSTD = 0.1354$. It is

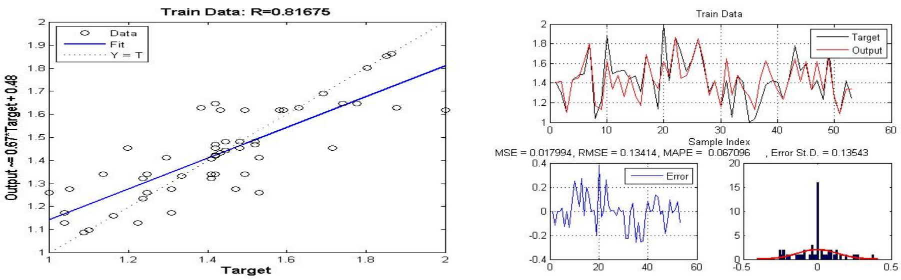
Table 3. Analysis results of proposed model for predicting chronic diseases.

Metrics	Asthma diseases	heart diseases	diabetics diseases	kidney diseases
MSE	0.0331	0.0180	0.0142	0.6910
RMSE	0.1819	0.1341	0.1193	0.8312
MAE	0.7874	0.0671	0.0577	0.1739
ErrorSTD	0.1836	0.1354	0.1205	0.8392



(a) Regression plot. (b) Time series plot.

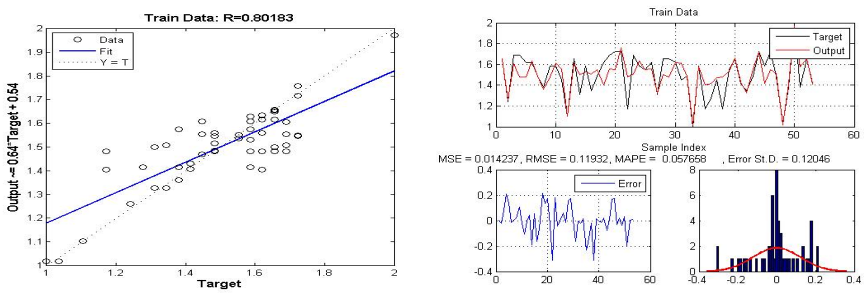
Fig. 2. Performance time series of ANFIS mode by using asthma diseases data.



(a) Regression plot. (b) Time series plot.

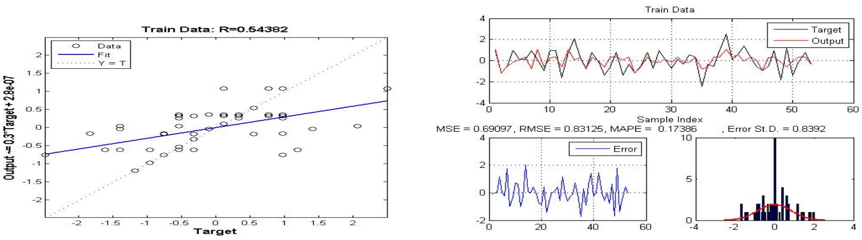
Fig. 3. Performance time series of ANFIS mode by using heart diseases data.

observed that the prediction of errors are very less as compared with asthma diseases and this indicates that heart clinical data and internet search data have the solidest correlation. The ErroSTD is very less as shown in Fig. 2. In Fig. 3 illustrates the percentage of correlation between the actual data and prediction heart data obtained from the soft computing model. The correlation R = 0.816 points out most prediction data is near to line regression. The predicted results of diabetics used by the proposed model are much better as compared with other chronic diseases. The proposed model has achieved 0.0142, 0.1193, 0.0577 and 0.1205 in terms of evaluation metrics to MSE, RMSE, MAE and ErroSTD. It is observed that the prediction of errors is very less in the time series plot as shown in Fig. 4. The percentage of correlation between output and actual diabetics data is R = 0.801, this is concluded the correlation between the registered data in official health and web search data searching by patients. Moreover, the prediction of results of kidney diseases used by the proposed model is satisfied. It is observed that the errors are more as compared with another chronic disease data. The prediction results are MSE =



(a) Regression plot. (b) Time series plot.

Fig. 4. Performance time series of ANFIS mode by using diabetics diseases data.



(a) Regression plot. (b) Time series plot.

Fig. 5. Performance time series of ANFIS mode by using kidney diseases data.

0.6910, RMSE = 0.8312, MAE = 0.1739 and ErroSTD = 0.8392. Fig. 5 shows the performance of the proposed model for predicting kidney data used by web search activities. It is also observed that the proposed model is more. The estimation of the proposed model for CDC – reported non-communication disease visits has given the possibility for improving the healthcare system for prediction non-communication diseases. The proposed methodology has strange predictive used by soft computing ANFIS. The results have indicated that the web search queries data are proxy to estimate clinical data. Consequently, the proposed model can improve the health level of the population at some time government. The web search queries are represented population-level trends in risk for the USA. The proposed system is examined that web terms are used as a proxy for non-communicable disease risk behavior at USA citizens. When internet search has become ubiquities through the smart health system, the health system will be more interactive with Google to process these words and exact useful information. The level of useful information is provided the opportunity to understand the effeteness heterogeneous policy change in different countries. This research work made a potential power for developing a smart surveillance system that detects chronic diseases.

In fact, the USA has annual surveys of non-communication diseases but another country does not have such kind of this system, we have used USA country data for training the proposed system. In further work, we try to develop the annual surveys of non-communicable diseases in Saud Arabia for making the healthcare system smarter. We have judged that the proposed system will be more applicable to developing the surveillance system in Saudi Arabia. The results have been obtained in this study recommend that web search is provided a proxy for non-communicable disease risk.

To validate and test this research work results, the experiment indicates that there are strong relationships between the clinical and hospital and search web that submits in the same period. USA populate ion suffering from spreading chronic diseases as WHO reported. Developing a surveillance system is a master required to stop spread these epidemics. As a result, the healthcare is very important for growing the economy of the country. There is an existence of surveillance system for collecting and reporting the data submitted into the official health. The prediction analysis approved that there is a temporal correlation between the Google trend and chronic disease risk. The data related to the surveillance system RSPS is correlated with Google trend as it is observed in the strongest relation. Finally, it is observed that with the increasing search on the web about information health will open the large doors to WHO for improving healthcare. Hence, the WHO with using web search main advantage of using search quarries can obtain useful information for early detecting and prediction risk. The information from web search informs the government due the data of CDC is later for one week.

5. CONCLUSION

The chronic disease is one of the biggest diseases faced societies in over the world. For this reason, the proposed research is conducted to avoid and minimize the spread of the disease prediction and the chronic diseases by using the new system. The new novel is to predict chronic diseases by using web search activities that have been submitted and searched by patients within sociality. The present study is the focus of the USA sociality. The main objective of the proposed system is to design an advance model to handle and predict non-communicable diseases in order to help official health and control the spreading diseases. The non-linear soft computing was applied to predict chronic diseases. The standard data is gathered from the CDC. The internet search pattern is obtained from the Google trend. Four evaluation metrics are employed to test the proposed system. From the experimental analysis, it is investigated that we can design system that predicts chronic diseases namely with using search terms. It is concluded that there are relations between the clinical data that registered in public health and search queries that are submitted by the population.

ACKNOWLEDGEMENT

The authors acknowledge the Deanship of Scientific Research at King Faisal University for the financial support under Nasher Track (Grant No. 186164).

REFERNCES

1. R. Nagar1, Q. Yuan, C. C. Freifeld, M. Santillana1, A. Nojima1, R. Chunara1, and J. S. Brownstein, "A case study of the New York City 2012-2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives," *Journal of Medical Internet Research*, Vol. 16, 2014 p. e236. 10.2196/jmir.3416.
2. M. A. Konerman, L. A. Beste, T. Van, and B. Liu, "Machine learning models to predict disease progression among veterans with hepatitis C virus," *PLoS One*, Vol. 14,

2019, p. e0208141.
3. A. Valdivia, J. López-Alcalde, M. Vicente, M. Pichiule, M. Ruiz, and M. Ordobas, "Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks – results for 2009-10," *Eurosurveillance*, Vol. 15, 2010.
4. J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, Vol. 457, 2009, pp. 1012-1014.
5. S. Cho *et al.*, "Correction: Correlation between national influenza surveillance data and Google trends in South Korea," *PLoS ONE*, Vol. 9, 2014, p. e81422.
6. D. Seo *et al.*, "Cumulative query method for influenza surveillance using search engine data," *Journal of Medical Internet Research*, Vol. 16, 2014, p. e289.
7. D. Broniatowski, M. Paul, and M. Dredze, "National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic," *PLoS ONE*, Vol. 8, 2019, p. e83672.
8. R. Nagar *et al.*, "A case study of the New York city 2012-2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives," *Journal of Medical Internet Research*, Vol. 16, 2014, p. e236.
9. J. Santos and S. Matos, "Analyzing twitter and web queries for flu trend prediction," *Theoretical Biology and Medical Modelling*, Vol. 11, 2014.
10. N. Generous, G. Fairchild, A. Deshpande, S. D. Valle, and R. Priedhorsky, "Global disease monitoring and forecasting with wikipedia," *PLoS Computational Biology*, Vol. 10, 2019.
11. D. McIver and J. Brownstein, "Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time," *PLoS Computational Biology*, Vol. 10, 2014, p. e1003581.
12. P. Markey and C. Markey, "Annual variation in Internet keyword searches: Linking dieting interest to obesity and negative health outcomes," *Journal of Health Psychology*, Vol. 18, 2012, pp. 875-886.
13. P. Chowdhury *et al.*, "Surveillance for certain health behaviors, chronic diseases, and conditions, access to health care, and use of preventive health services among states and selected local areas behavioral risk factor surveillance system, United States, 2012," *MMWR*, *Surveillance Summaries*, Vol. 65, 2016, pp. 1-142.
14. M. Santillana, E. Nsoesie, S. Mekaru, D. Scales, and J. Brownstein, "Using clinicians' search query data to monitor influenza epidemics," *Clinical Infectious Diseases*, Vol. 15, 2014, 2019, pp. 1446-1450.
15. M. Kang, H. Zhong, J. He, S. Rutherford, and F. Yang, "Using Google trends for influenza surveillance in South China," *PLoS ONE*, Vol. 8, 2019, p. e55205.
16. Q. Yuan, E. Nsoesie, B. Lv, G. Peng, R. Chunara, and J. Brownstein, "Monitoring influenza epidemics in China with search query from Baidu," *PLoS One*, Vol. 30, 2013, p. e64323.
17. G. Milinovich, S. Avril, A. Clements, J. Brownstein, S. Tong, and W. Hu, "Using internet search queries for infectious disease surveillance: screening diseases for suitability," *BMC Infectious Diseases*, Vol. 14, 2014.
18. S. Towers, *et al*., "Mass media and the contagion of fear: The case of Ebola in America," *PLoS One*, Vol. 10, 2015, p. e0129179.

19. 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, 2018, pp. 1293-1299.
20. D. Huang and J. Wang, "Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors," *Science of The Total Environment*, Vol. 612, 2018, pp. 1293-1299.

Theyazn H. H. Aldhyani was born in Al-Dhalea, Yemen in 1986. He received the B.S.c and M.S.c. and in 2017 he was awarded the Ph.D. degree in Computer Science and Information Technology from NMU University, India. He published several research papers in the field of Computer Science and Information Technology and participated in many national and international conferences. His research interests include machine learning, soft computing, big data, healthcare information, network security, biomedical image processing, network traffic and internet of things. Currently, he is working as an Assistant Professor in King Faisal University, Community College of Abqaiq-Kingdom of Saudi Arabia, he is Chairman of Research Center in Community College.

Ali Saleh Ali Alshebami was born in Sana'a City of Yemen in 1985. He received his B.B.A (IT) degree from Symbiosis International University and the MBA degree from Dr. BAM University, India in 2009 and 2011, respectively. The Ph.D. degree in Management Science (Microfinance field) was also received from the SRTM University, India. He published several research papers per- taining to microfinance industry, Financial services pricing, electronic government security and other fields. He also published a book on the Basics of Microfinance. He participated in many national and international conferences and offered various works shops and lectures on topics related to self development and management. He worked as Head of Tenders and International Relationship Department in Alshebami Group in Yemen for two years. He is currently working as an Assistant Professor in King Faisal University, Community College of Abqaiq-Kingdom of Saudi Arabia. He is also in charge of many committees in the college.

Mohammed Yahya Alzahrani has received his B.S.C in Computer Engineering from Albhaha Private College of Science and M.S.c in Information Technology from Heriot Watt University, Edingburgh, UK. He then obtained his Ph.D. in Computer Science from Heriot Watt University, Edingburgh, UK. Currently he is working as the Dean of College of Computer Sciences and Information Technology at Abaha University in KSA. His research interests include in data mining, model checking and cyber security.