Estimating Diabetic cases in KSA through search trends and Creating Cyber Diabetic Community

Yasir Javed[1, 2]
[1]FIT, UNIMAS, Malaysia
[2] Prince Megrin Data Center,
CCIS, PSU,Riyadh, KSA
yjaved@psu.edu.sa

Adnan Shahid Khan
FIT, UNIMAS, Malaysia
skadnan@unimas.my

Basit Qureshi
Prince Megrin Data Center,
College of Computer and
Information Sciences, PSU
qureshi@psu.edu.sa

Junaid Chaudhry
Department of Computer
Science, Innopolis University,
Kazan, Russia
j.chaudhry @ innopolis.ru

Abstract— **Saudi Arabia is fastest developing nation enjoying stability and high per capita income, thus highly influenced by urbanization inviting huge investments from international brands especially in food and clothing sector. Changes in life style have made Saudi society more prone towards disease like diabetes that is costing about 40% cases of the total population [1]. Higher diabetic cases have alarmed health care organizations (both government and private) in finding the exact number of diabetic cases in extremely timely manner. Creating a unified system among hospitals, laboratories and other health care organizations is time consuming and expensive inviting researchers to look for other options. Due to improvement in community awareness among stakeholders (patients, care taker, health researcher and others), has provided an opportunity to get a real time estimate about total number of patients and getting to know about patient problems etc. This study tends to create a diabetic prediction system that will gather information from multiple sources (news, health care records, social media, news feeds, search trends and tweets) in multiple languages (Arabic , English and French) to answer two questions (1) Can online search trends and tweets be related to exact number of diabetes patients (2) Can we extract common or new symptoms for diabetes cases from these trends (3) providing a predictive picture to health care professionals and managers for creating in-time policies to avoid epidemic.(4) Finding relationship in between diabetes related search terms and diabetic cases.**

This study reveals that real data figures are 85% correlated to search trend thus providing a cogent proof that both internet usage and real data figures can be related. It was also observed that search trends commonly symbolize common symptoms or disease name. A cyber diabetic community can be created that can be targeted by government agencies or health organization as to create awareness about diabetes. While usage of system by community will also help in better diagnosis from search trends and hospital information.

*Index Terms—*Prediction of Diabetes, Google Trends, Correlation of healthcare data, Cyber diabetic community

I. INTRODUCTION

World is beautiful and an enjoyable place where we can live happily and healthy, but fights, political tension and instability have causes lots of death but most number of deaths are caused by diseases, sickness, stress, mental disorders and chronic disease. It has been evident that most number of deaths that are caused due to any of these both cases belongs to poor nations where disease control is not there. Chronic disease are one of major life taking. Chronic disease are of two types, one where diseases are spread due to hygienic conditions or health services like hypertension, heart diseases and cancer but there are chronic disease that are mostly related to developed countries like diabetes, obesity and cardiovascular diseases. In overall all countries in world chronic disease cost 17% of deaths per year according to WHO [1]. As mentioned there are couple of chronic disease like AIDS/HIV, MERS, Cholera, Cancer, Cardiovascular and Diabetes. Diabetes is costing Chronic diseases holds a major share of global disease burden for overall population, it also accounts for major deaths in almost all countries i.e. almost up to 17% according to WHO [1]. In 2013, diabetes costed 51 Million deaths [2].

Diabetes is a long term diseases that is mostly incurable holding a major share of patients worldwide. It requires a regular medication on daily basis and requiring a proper monitored diet. Diabetes is caused due to endocrine hormonal disorder whose function is to convert the food that is being intake by converting it to glucose that is used as energy source. There are numerous symptoms for diabetes but not limited to these, symptoms includes increased frequency of urine, unexpected loss of weight, hunger or thirst, healing process is very slow or in some case no healing and mood swings. [3]
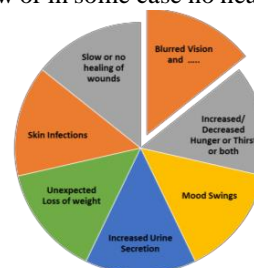

*F*igure 1: Showing some possible symptoms of diabetes.

There are three major types of diabetes Type 1, Type 2 and gestational diabetes. A patient is classified of having type1 diabetes when the body is unable to make insulin. Type 1 is also called as IDDM (insulin-dependent diabetes mellitus). There is no cure or prevention of this type and it accumulated the 10% of total patients in the world [4]. Another major type of diabetes is Type 2, where body shows no response to insulin so the patient are referred as NIDDM (Non-insulin dependent mellitus). Type 2 diabetes is usually caused by modified life style as well as some genetic cases. Type2 accumulates about 85 to 90% of total cases in the world. Third major type is gestational diabetes that occurs during pregnancy but goes away once baby is delivered. There are almost 5 to 10% cases in world reporting this type [5].

Diabetes causes a lots of deaths or disabilities every year causing only in 2014, 4.9 million deaths. Till 2014, the prevalence of diabetes in world was 8.3% that makes 837 million people diagnosed with diabetes, while still 43% people are untested [6]. This constitutes a very high number and thus it needs to be addressed.

Diabetes prevalence is very high in Middle east like Kingdom of Saudi Arabia 23.8 % while Bahrain 21.9%, Qatar and UAE 19.8 and 19 respectively.[6] as shown in Table 1. While the table also show the Gross National Income per capita (GNI). This table only includes some of the countries that have very high GNI in order to evaluate the reasons of diabetes.

Country Name	Diabetes Prevalence in 2014	GNI in USD in 2013/14
Saudi Arabia	23.8%	$25,140
Bahrain	21.9%	$21,050
Qatar	19.8%	$94,410
UAE	19%	$45,200
Oman	14.5	$16,870

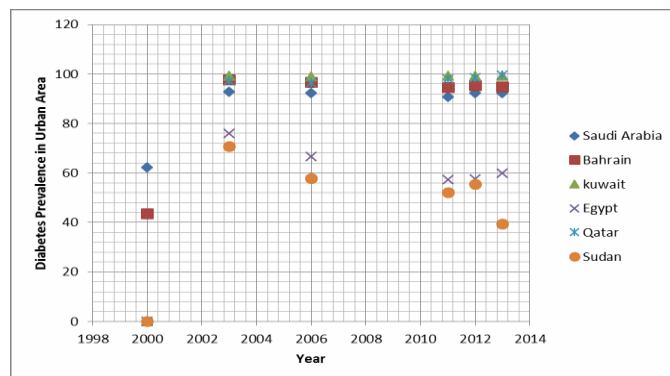Table 1: Showing the prevalence of diabetes in Middle East in 2014 [6], [7]

Figure 1-A: Showing the prevalence of diabetes in Urban Areas in Gulf Areas

High prevalence shows that these cases has to looked into and cannot be ignored, this study will currently look into only KSA that is a fastest developed nation, enjoying stability but having the most high prevalence of diabetes. KSA government is subsidizing around 98K SAR to 180K SAR for each patient in hospital that is resulting in lots of expenses in patients [8]. Also depicted in *Figure 1-A*, the prevalence is much higher is urban areas of KSA, Bahrain and Kuwait.

As shown in *Table 1* KSA per capita GNI is above 25K USD responding to a making KSA into list of rich nations. Here mostly due to lifestyle people tend to spend most on their cars, decorations, travelling and shopping but most of them spends on food that is 118 billion USD by 2016 [9]. According, to government and food agencies of KSA annually the food wastage is 1.3 Billion ton food [10] that includes sandwiches, fast food, rice and other fired items. Presence of hundreds of food chains also demonstrate the type of food that KSA like. Numbers of cars increases every year and so are the electronic gadgets, Shoppers are increasing every year and number of people getting loans after high capita income also depicts the amount of money splendid. Petrol is subsidized and is available at very low cost making people to travel mostly on cars or other vehicles even when distance can be done via foot or bicycle. Golf carts are available inside big malls avoiding physical exertion that can burn some calories. Table 2 shows the food consumption in GCC for the year 2012 that shows a major per capita consumption of food in which Kuwait and KSA holds the major consumption of food.

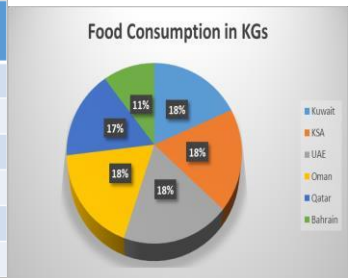| Country Name	Food Consumption in KGs	
Kuwait	867.5	
KSA	866.5	
UAE	862.2	
Oman	845.7	
Qatar	775.6	
Bahrain	491.1	

Table 2: Showing the Food Consumption of GCC and figure also showing the share of food consumption in GCC (statistics obtained from source: IMF, Arab agriculture statistics 2012, Alpen Capital)

Each thing representing the KSA is fully urbanized. But Urbanization has an effect, like increased number of diabetes cases about 23.8 % and in adults the number of diabetic patients are more. Urbanization has resulted in life style change, eating habits and physical works and so is increase in the chronic diseases that are related to urbanization. Increase in diabetic cases in most developed countries is also an indicator of this.

Increasing diabetic patients in KSA require awareness in terms of education, life style changes, eating habit changes and proper medical attention. KSA is striving hard to raise the awareness through Ministry of Health, diabetic societies, conferences, seminars, hospital awareness programs and media awareness campaigns. These campaigns are fruitful to raise the awareness but still rise in patients required to government

monitor the number of patients having diabetes. Only data that is available through is from hospital records that are not connected. Studies conducted are also on only few patients at single hospital made it very difficult to find out exact number of patients. In order to find out the exact number of patients there are numerous ways, a KSA wide survey to find out real number of patients, getting all records from hospitals, interconnecting hospital to a system where all records are loaded but each method requires a huge amount of time and money to be splendid to make a system.

Internet usage is continuously increasing in KSA as shown in Figure 2.
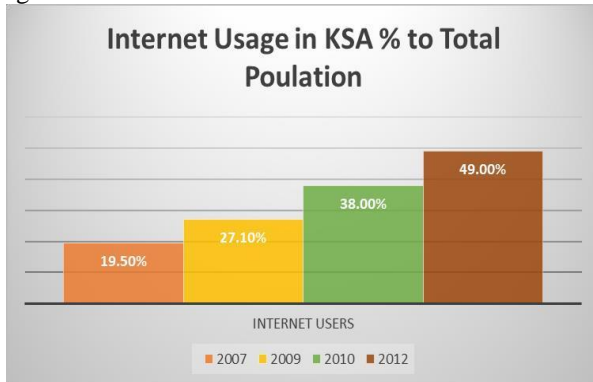

F*igure 2: Showing the internet usage per total population in KSA as per ITU*

Internet usage statistics shows that till 2012 49% people were using internet in KSA as per ITU, while looking at previous years that is 19.50% in 2007 to 49% shows a major number of users currently. Internet usage increase and literacy programs also depicts that people will use internet for finding disease related information. This information can in be form of symptoms, types of disease, cure, medications and its famous doctors working in this area, for diabetes same statements holds. The stakeholder that will look for information will be patients, their care takers, hospital staff and research staff. Information finding on internet is done through usually search engines. There are couple of famous search engines like Google, Bing and Yahoo. But most common is Google in case of search engine used to find an information. In KSA, about 58% Google is used as searching tool. Even from Jan to Oct there were 13324 Trillion searches [12].
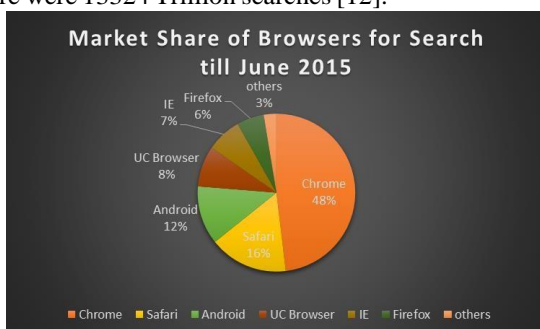

Figure 3: Showing the browser usage for June 2015 [13]

From *Figure 3*, it is evident that chrome is mostly used but this doesn't means google is used as search engine, but still

Android also refers to google thus constituting 60% of the people total as a browser.

This study tends to correlate the issue of finding the real diabetes cases with number of searches made, in order to find the easiest and almost free solution in finding the real number of test cases. KSA is multinational state where Saudi are about 18 Million and 8.9 Million Expats [14]. Arabic is a national language and is spoken most widely, also the Arabian people like expats from Sudan, Yemen, Egypt and Emirates also speak Arabic. Second most common language is English that is known by most people like Arabs and Expats due to high literacy rate. Moreover mostly the research is done in English language, so most of doctors and research staff look into the web for information in English. Third most common language for search is French due to high literacy rate and French being taught as second language in schools around KSA. There are other languages that is mostly spoken like URDU, Bangladeshi, Tamil but this study will only focus on first three languages that are Arabic, English and French with respect to diabetes.

This paper is divided into following sections firstly a research object of study will be defined then research methodology will be presented. Results will be presented in next section with discussion about results and in end conclusion will be provided.

II. RESEARCH OBJECTIVES

This study will be focused on following questions that are as follows

1. Can online search trends and tweets be related to exact number of diabetes patients?
2. Can we extract common or new symptoms for diabetes cases from these trends?
3. Providing a predictive picture to health care professionals and managers for creating in-time policies to avoid epidemic?
4. Finding relationship in between diabetes related search terms and diabetic cases.

III. RESEARCH METHODOLOGY

In order to cover the research objective, this study will cover work in two aspects. First aspect is finding the real number of diabetes case that will be base for our study and will be used as bench mark too. Second aspect is to find the search trends that has occurred in KSA for a span of ten years.

For first aspect, this study will use research papers that publishes the result collected from hospitals, Diabetes Atlas and from International Diabetic Federation. For the second aspect Google search trends will be used to find each term usage from the year 2004 to October 2015. Google search trends have been used by Carnerio et al [15], Palet et al [16] and Chalet et al [17]. Google Search trends [18] [20] can provide the search per term, per month as per country and its region.

Initially this study collected the actual number of patient records from the IDF Atlas and then search terms were

shortlisted that focuses on disease its self, its treatments and its medication, For each language separate terms were selected as shown in Table 3. Most focusing terms will be For Arabic and English, search terms includes diabetes, insulin as its medicine and finding the symptoms in all three languages.

English	Diabetes, Diabetes Mellitus, Diabetes 2, diabetic , diabetes symptoms and Insulin
Arabic	السكري مرض, السكري مرض علاج السكر, السكري مرض اعراض,والسكر علاجand الأنسولين
French	Diabète, diabète sucré, insuline , symptômes du diabète and diabétique

Table 3: Showing the selected search terms used in different languages for finding information related diabetes.

IV. ANALYSIS AND DISCUSSION

Average age for KSA is above 70 years as per WHO so this study focuses on age group of 20-79 years. The data set are obtained from IDF world Atlas for diabetes for the year if 2003, 2006, 2009, 2011, 2012 and 2013 as shown in *Table 4*. The table divides the population according to Urban and Rural group. By looking at figure it is evident that number of people living in Urban areas. It can be due to couple of reason like people shifting to cities, or expansion of cities and including nearby villages or creation of new cities. In case of KSA all three holds.

		Number of people with DM (000's) in the 20-79 age-group		
		Total Population	# of Diabetic Patients	Rural	Urban
Years	2000	10,581	996.7	376.3	620.4
	2003	10,544	992.2	70.5	921.7
	2006	13,730	1,854.90	140.3	1,714.60
	2009	15,187	2,065.04	197.9	1,867.14
	2011	17,024	2,759.56	257.659	2501
	2012	17,582	3,415	265.732	3148.78
	2013	18,057	3,651	275.909	3374.98

Table 4: Number of Diabetic Patients in Multiple Years with Urban and Rural Distribution using IDF data extracted from [3]

As per *Table 4*, each year number of diabetic patients are increasing and so are patients in Urban areas, it may also depict that due to urbanization the diabetes is more prevalent in urban areas than rural area.

Second part of this study is to find the search trends conducted in KSA for the related terms mentioned Table 3.

Figure 4 shows the initial search terms for السكر, Diabète. Orange line for **Diabète** is not shown as it doesn't constitute to total search trends, while blue lines that for **diabetes** the search trends are consistent almost that makes it evident that people are searching about diabetes, while السكر trends in redlines refers to high number of search trends and they are increasing since start. These things refers to increased awareness and literacy rates among Arabic users, while also

meaning increased diabetic cases when correlated with number of diabetic patients as defined in *Table 4*. Google search trends also shows that which city or region is searching more like shown in *Figure 5*.
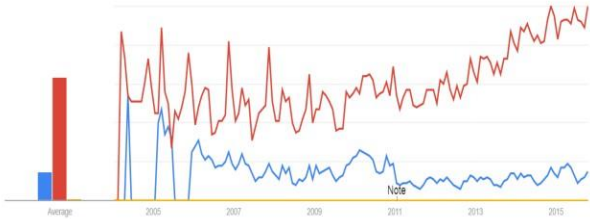

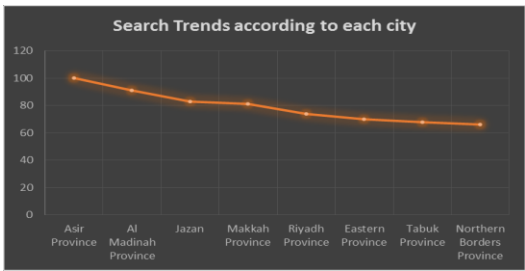Figure 4: Showing the search trends from google search trends for diabetes, diabète and السكر

Figure 5: Showing the search trends from google search trends for each city in KSA from 2004 to October 2015.

Looking at Figure 5, as when the correlation is proved this figure can provide means to find the number of diabetic cases in each city which can help in finding the number of diabetic patients, planning required to curtain the disease and other actions at right time before the disease is spread.
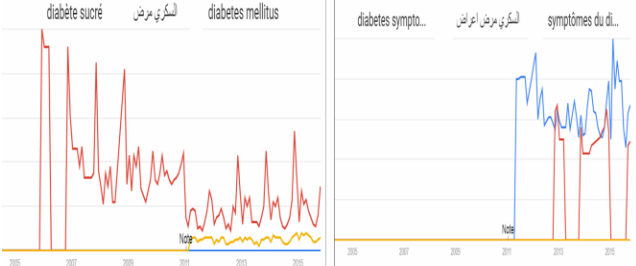

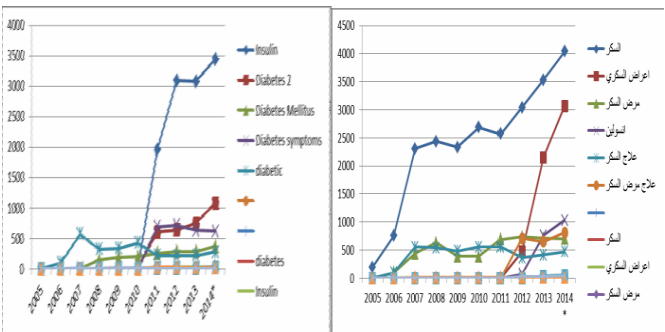Figure 6: Showing the search trends from google search trends for numerous search terms.

Figure 7: Showing the search trends from google search trends for Arabic and English terms till year 2014 [20]

It is evident from Figure 6 that after 2011 the search trends have increased correlating to sense of awareness of diabetes as well as increased number of diabetic cases in persons who have easy access to internet and who use internet for the purpose of care or patients and want to know about disease. While looking at Figure 7 it is seen that most number of searches are for the disease name and its types, while in Arabic the search also includes the symptoms. The figure doesn't include the French terms as they don't make a much impact on search trends and thus can be ignored.

Both cases results are now calculated that means real number of diabetes cases and the search trends are also observed, now both of them have to be correlated in order to find out how diabetic cases are related to search trends. Firstly, all of the search terms have been taken sum $\sum_T$ where T are search over the total span of time which is our case is period from 2004 to October 2015. Now, all type of searches made in each month will be calculated called as $\Re_S$, where S is number of searches in each month while $\Re$ is the total sum. Percentage of diabetic patients β will be now calculated as shown in *Equation (1)*

$$\beta = Percentage(\sum_T, \Re_S) \qquad \textit{Equation 1}$$

Conversion has to be done as Google search trends are in percentage of total search trends.

The total population each year is represented by P_i where *i* is the *ith* year. The population relation to total number of searchers $\prod$ can be made by dividing total number of searches with population as the searches are done by population.

$$\prod = \sum_T / P_i \qquad \textit{Equation 2}$$

But the resultant from *Equation 2* this will not be a number compared to search trends so a constant ℓ is multiplied to make a comparable mapping between search terms and diabetic patient in overall population. So modifying *Equation 2* we will have following

$$\prod = \left(\frac{\sum_T}{P_i}\right) * \ell \qquad \textit{Equation 1}$$

Equation 3 will present a correlation between search trends and number of diabetes cases as obtained in Table 4
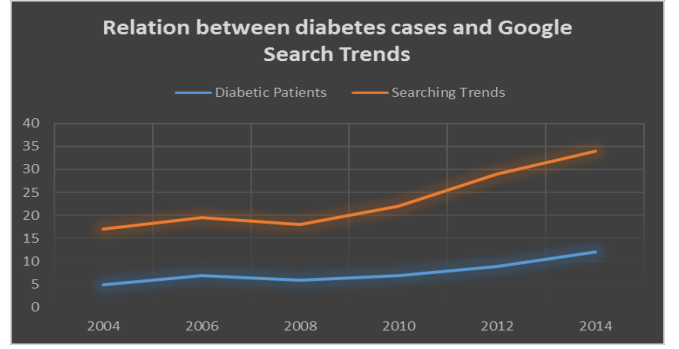
.

Figure 8: Correlating of search terms with diabetes cases in KSA

As seen in Figure 8 the trends of both diabetic patients and search trends can be correlated and is visibly depicted that both carries the same meaning. There is difference between the numbers as seen the search trends are higher but if curves are matched, it is up to 85% same curve.

To validate the equations a bi-nominal equation can be used to find out the real correlation between both cases. Bi-nominal equation is given as follows. Where S are search trends or can be used for diabetic patients. And a and b are constants.

$$\gamma = aS^2 - bS + \ell \qquad \textit{Equation 4}$$

Substituting the values in Equation 4 results in 0.84 that shows a high correlation between both the cases.

Once this is matched, it also means that results obtained in Figure 5 can also be used to find out the estimated cases in cities or regions which is still hard to find out and require man power and money.

V. FUTURE WORK

This also gives an idea of making the system more viable by introduction of other search trends tools like Facebook, twitter and line but not limited to these three. There will a complete collection model that will collect results from search trends, Facebook, twitter, internet data and real data from all the hospital database and will be able to store the data. System then on that data can run data mining and data analytics algorithm in order to correlated between the data and figuring out new symptoms, new cures or at least some cure. The predictive behavior of the system can help all the stakeholders and will save cost, time and will be improved over period of time. The system should support privacy and records should be secure in order to not allow unauthorized person to use the basic information of patient. Thus in order to achieve privacy, users of system will get anonymous data, while a patient and his physician will be able to see the patient data and his related trends.

KSA if seen through internet usage, is becoming a real cyber community. Thus needs of cyber agents are required to be established to promote the awareness among the community. As diabetes, if genetic cannot be avoided but

awareness among fully active cyber community will yield a healthy life. Patient searches for diabetes related information can be very useful as these searches can provide through the information what patient is going through. It will solve the issue of language barrier or in ability to explain what patient is going through.

VI. Conclusion

This paper primarily focuses on finding out the relationship of Google Search trends to real diabetic cases. Diabetes is referred as silent killer as there is no cure to this disease, moreover in world still 43% of people are untested. Diabetes is most common in developed nations like Middle East Area. In KSA, diabetes is 23.8% prevalent while there will be still patients like younger people, child's or rural areas people whose disease is never tested or diagnosed. Also to find out the real number of patients is hard, time consuming and expensive procedure. This paper uses internet search trends as a mean, as KSA is literate society and penetration of internet is too much that is 49%. This paper addresses basic points like what is diabetes and how it is present in KSA and how much internet is prevalent in KSA? While this paper also addresses basic four (4) questions (a) Can online search trends and tweets be related to exact number of diabetes patients? (2.) Can we extract common or new symptoms for diabetes cases from these trends? (3.) Providing a predictive picture to health care professionals and managers for creating in-time policies to avoid epidemic? (4) Finding relationship in between diabetes related search terms and diabetic cases. We used *Diabetes, Diabetes Mellitus, and Diabetes 2, diabetic, diabetes symptoms and Insulin* as English search terms. *Diabète, d*iabète sucré, insuline , symptômes du diabète and diabétique are used to search about diabetes in French and

الأنسولينandالسكر علاج , السكري مرض اعراض ,السكري مرض علاج ,السكر ,السكري مرض

It is observed that about 85% of values are correlated to each other. The result provide a conclusive evidence that Google search trends can be used to estimate the real diabetes cases. It can also allow to find out in which area diabetes is more prevalent. Timely results can yield in better management of disease and help in case of required help. Increase search trends also depicts the awareness programs can be a useful told to make this approach successful.

In future, we plan to create a Cloud based cyber diabetic society system, which can handle the streams of data coming from various sources like twitter, Facebook and search trends. The actual data will also be made available through feeds from hospitals and clinics. Once that data is in the system, data analytics algorithm can used to analyses the data and generate results in perspective of patients, their problems, care takers and health physicians. The system will provide government and health agencies a timely forecast to take the preventive measures. Moreover the system than can be extended to multiple type of disease resulting in a complete medical system.

REFERENCES

[1] Alberti, Kurt George Matthew Mayer, and PZ ft Zimmet. "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation." Diabetic medicine 15.7 (1998): 539-553.

[2] Diabetes Australia. Third largest diabetes society. https://www.diabetesaustralia.com.au/ as of October 2015.

[3] Diabetes Atlas, 6th ed. Brussels, Belgium: International Diabetes Federation.

[4] World Bank Data, http://data.worldbank.org/country as of Oct 2015

[5] http://www.arabnews.com/diabetes-among-saudis-major-issue Accessed October 2015

[6] World Health Organization, ed. World health statistics 2015. World Health Organization, 2015.

[7] Diabetes Atlas, 6th ed. Brussels, Belgium: International Diabetes Federation.

[8] Al-Nozha, Mansour M., et al. "Diabetes mellitus in Saudi Arabia." Saudi medical journal 25.11 (2004): 1603-1610.

[9] http://www.arabnews.com/news/economy/611661 as of October 2015.

[10] http://www.arabnews.com/saudi-arabia/news/722026 as of October, 2015.

[11] Fung, Teresa T., et al. "Dietary patterns, meat intake, and the risk of type 2 diabetes in women." Archives of internal medicine 164.20 (2004): 2235-2240.

[12] Search Engine usage statistics report by Global Stats Counter- US Till October 2015.

[13] Internet usage, broadband and telecommunications report, Saudi Arabia, IWS (June 2012)

[14] Center Department of Statistics and Information, Government of KSA. For Year 2010

[15] Carneiro, Herman Anthony, and Eleftherios Mylonakis. "Google trends: a web-based tool for real-time surveillance of disease outbreaks." Clinical infectious diseases 49.10 (2009): 1557-1564.

[16] Pelat, Camille, et al. "More diseases tracked by using Google Trends. "Emerging infectious diseases 15.8 (2009): 1327.

[17] Choi, Hyunyoung, and Hal Varian. "Predicting the present with google trends. "Economic Record 88.s1 (2012): 2-9.

[18] Google Trends: A Web Tool Used for Collecting Data www.google.com/trends as of October 2015.

[19] Zhou, Xichuan, et al. "Monitoring epidemic alert levels by analyzing internet search volume." Biomedical Engineering, IEEE Transactions on 60.2 (2013): 446-452.

[20] Y. javed et all "Correlating new diabetes diagnosis cases for Saudi Population with Google Search Trends" Poster Paper in SDMA 2014, PSU