Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Hierarchical K-Means Clustering Algorithm for an E-Care of Diabetes Mellitus

T.Karthikeyan,
Associate Professor,
Department of Computer Science, India

R.Ragavan,
Associate Professor,
Department of Bio Chemistry, India

K.Vembandasamy
Ph.D Research Scholar &
Assistant Professor, Department of Computer Science,
PSG College of Arts and Science. Coimbatore-14, India

Abstract - The Internet-based applications are being developed to support patients in self-managing their diabetes by giving them the opportunity to e-mail with their caregiver, to monitor their health, and to enhance self-care by means of online education. The major issue in people nowadays is the lack of awareness about eating habits. Prevalence of diabetes (mellitus) in our country has steadily increased for the reason. Today there is an increase in interest for setting up medical system that can screen a large number of people for life threatening disease such as Cardio Vascular Disease (CVD), retinal disorder in diabetic patients. The main aim of this study was to assess the value of the Internet for supporting self-care of patients with diabetes. We investigated user motivations and experiences with the Internet-based application. We focused on implementation, user-friendliness, and the process of care delivery via the application. The main objective of this study was to assess whether self-efficacy (SE) could function as a moderator of the effect of a tailored Internet-based intervention aimed at increasing self-reported diabetes self-care behaviors. The ultimate goal of diabetes care management is to optimize self-care management in order to reduce the mortality, morbidity, and health care costs. Optimal diabetes management involves considerable behavioral modification. In order to promote healthy behavior, effective health communication is essential. The introduction of the Internet into clinical practice as an information-sharing and communication medium has brought about many opportunities for providing immediate, transactional feedback on lifestyle modification, next to regular ways health care delivery. As a result, interactive and responsive e-health applications with continuous self-monitoring, feedback, and information exchange are more and more being developed. So, In this paper K-means method is used. K-means is implemented using standard Euclidean distance metric, which is usually insufficient in forming the clusters. The tool used in this work is MAT Lab.

Keywords- diabetes management, self-care, Treatments, e-health applications, K-means.

I. Introduction

Data Mining (DM) is one of the most useful methods for exploring large data sets. Clustering, a special area of Data mining is one of the most commonly used methods for discovering the hidden structure of the considered data set. The main goal of clustering is to divide objects into well-separated groups in a way that objects lying in the same group are more similar to each other than to objects in other groups. Clustering can be used to quantize the available data, to extract a set of cluster prototypes for the compact representation of the data set, to select the relevant features, to segment the data set into homogenous subsets, and to initialize regression and classification models. Clustering as an important tool to explore the hidden structures of modern large databases has been extensively studied and many algorithms have been proposed in the literature. Because of the huge variety of problems and data distributions, different techniques such as hierarchical, partitioned, density-based and model-based approaches have been developed and no particular technique is completely satisfactory for all cases. There are quite a good number of algorithms that are proposed under each category and applied in many domains that depend upon the nature of algorithms. As the applications of data mining in clustering increases day by day, the scope and necessity of finding smart and efficient clustering algorithms is also increased.

A thorough understanding of the genes is based on upon having adequate information about the proteins. Solving the protein related problem has become one of the most important challenges in bioinformatics. In bioinformatics, number of protein sequences is more than half million, and it is necessary to find meaningful partition of them in order to detect their functions. The main objective of the unsupervised learning is to find a natural grouping or meaningful partition using a distance function. The method which can enhance the structural recognition, classification and interpretation of proteins will be advantageous. Many methods have been adopted to solve such bioinformatics problem. Many methods are currently available for the clustering of protein sequences into families and most of them can be categorized in three major groups: hierarchical, graph–based and partitioning methods. Among these various methods,

most are based on hierarchical or graph-based techniques and they were successfully established. Our hierarchical k means clustering algorithm is useful and efficient method in the collective study of protein subset. Collective analysis of the proteins at the time when they are in numerous numbers and when one cannot study them individually may be very useful. Moreover there may be interesting patterns in each of the protein collection that may escape from attention when studied individually. Protein clustering is a method which can be very useful in the recognition of biomarkers and helping in their classification. The main objective of the k means clustering algorithm is to help classification and prediction of the biological functions as well as recognition of new interpretation patterns among them. Among these the most important ones include the protein sequences related to diabetes. Since our approach is combining the features of hierarchical and graph-based clustering, the centre of each cluster includes the protein cluster information that enhances the rapid analysis of the proteins. In fact, the number of protein sequences available now is very important and meaningful clusters to improve the classifications' quality and reduce the computation time compared to the available tools.

Improving care for patients with diabetes mellitus has become a priority for health plans, payers, and patients in the Netherlands and worldwide. The ultimate goal of diabetes care management is to optimize self-care management in order to reduce the mortality, morbidity, and health care costs. Optimal diabetes management

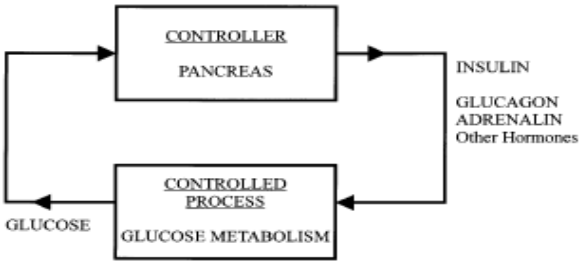

Fig. 1. A simple model of glucose: insulin interaction

The glucose: insulin loop is complemented by loops involving a range of other hormones, including glucagon and adrenalin. Whereas insulin has the principal effect of pulling down elevated glucose levels, the effects of the other hormones are the reverse; counteracting low levels of blood glucose, for instance by causing glycogen to be pulled out of its liver and muscle stores, thereby raising the level of plasma glucose. All these processes, in both physiological and pathophy siological states, are dynamic with time course profiles ranging from minutes to hours and beyond.

II – Related Works

A comparative survey design and the pretest data of a larger intervention study (Lee, Park, Park, & Kim, 2005) were used for the study. The survey was conducted from November 1, 2003 to June 30, 2004.

Participants were recruited from an endocrinology outpatient department in secondary (n = 1) and tertiary hospitals (n = 1) and public health centers (n = 7) in an urban city of South Korea. The selection criteria for the subjects were: (a) diagnosis of Type II Diabetes, (b) the absence of any severe physical disability that limited independent physical activities, and (c) possession of an intact cognitive ability. A total of 174 patients met the above-described criteria and written consents were obtained from all participants in the study.

For International Physical Activity Questionnaire (IPAQ; Ainsworth et al., 2000), Revised Summary of Diabetes Self-Care Activities Measure Scale (Revised SDSCA Scale;Toobert, Hampson, & Glasgow, 2002), and Diabetes management self-efficacy scale for patient with Type 2 Diabetes (SE-Type 2; Bijl, Poelgeest-Eeltink, & Shortridge-Baggett, 1999), the original developer of the instruments were contacted and get a permission to translate into Korean. Above instruments were translated into Korean and back translated by two Koreans who can freely use English and Korean Computational techniques in analyzing the dynamics of blood glucose and its control mechanisms has a long history. The first major wave of activity occurred in the 1960s, beginning with the cybernetic approach of Goldman, quickly followed by activity in a dozen or more North American centers. By the late 1970s and early 1980s dynamic modeling had achieved considerable maturity in terms of modelling methodology, incorporation of isomorphic physico-chemical effects and attention to potential clinical applicability.

With the advent of the microcomputer in the 1980s emphasis switched towards the wider application of information technology in the context of diabetes, though modeling regained its important position in the 1990s as an integral component of many computer-based decision support systems. The 1980s also marked the serious emergence of an additional strand of computer science in relation to diabetes, through the development of primitive expert systems and other manifestations of 'artificial intelligence' and knowledge-based technology.

In a classic paper published in 1988, Rodbard foresaw the potential role of computers in terms of: database systems in clinics and hospitals; microcomputers for use by physicians and patients; portable devices to provide recommendations regarding insulin dose; and memory equipped glucose meters.

Substantial progress has been made in all four areas over the ensuing decade. Examples in each case are to be found in other papers in this issue. One of the earliest hospital-based database and information management systems to find routine application in the diabetic clinic was the Diabeta system at St. Thomas' Hospital, London. Equally there are a number of locally developed systems to be found in primary health care.

A. Clustering.

Data mining functionalities like clustering and attribute oriented induction techniques have been employed to track the characteristics of the women suffering from diabetes. Information related to the study was obtained from National Institute of Diabetes, Digestive and Kidney Diseases. The evaluation of the characteristics is performed taking into account the data obtained from National Institute of Diabetes, Digestive and Kidney Diseases. Initially the data is grouped into clusters by using the clustering techniques. Once the data is grouped into clusters, the quality of the cluster needs to be identified as clusters without good quality are not helpful in effectively. Once the algorithm is identified the characteristics of the clusters (diabetes data) can be evaluated using the approach of Attribute Oriented Induction. In this approach the first step is to identify the number of distinct values of various attributes. After identifying such attributes, those attributes with maximum number of distinct values are removed. From the remaining attributes maximum and minimum values are identified and the items are grouped using the concept of set grouping by taking into account some threshold value.

In technical terms there have been expert systems to advise on patient management, computer data mining algorithms for insulin dosage adjustment, a range of analyzing diabetes as well as approaches of treatments upon optimal or adaptive control. Further detail of developments during the 1990s in all four areas can be found in a number of issues of the leading journals in the field that have been devoted to the themes of information technology and computing in diabetes.

III - K-Means Clustering Approach

In this paper, we use K-Means approach as follows:

The K-Means clustering algorithm is commonly used in computer vision as a form of image segmentation. The result of the segmentation are used to aid border detection and object recognition. In this paper, the standard Euclidean distance is usually insufficient in forming the clusters. It assigns each point to the cluster whose center is nearest. K-Means is one of the simplest unsupervised learning algorithm and follows partitioning method for clustering. Clustering based on K-Means is closely related to a number of other clustering and location problems.

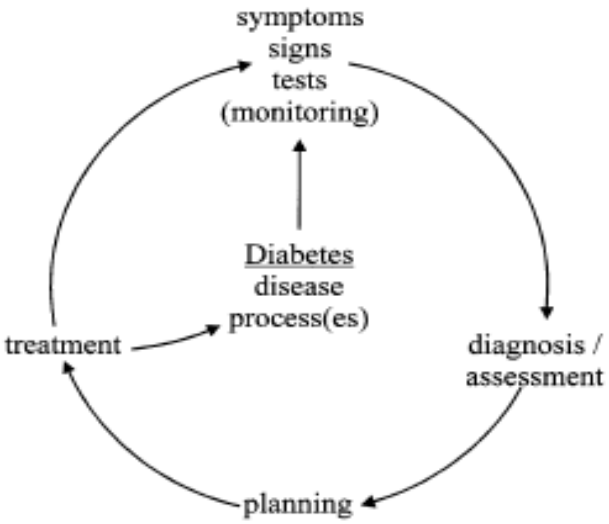

Fig. 2. A dynamic perspective of the clinical cycle of diagnosis: assessment, planning, treatment and monitoring of the diabetic patient

One of the most popular enabling person to discover for solving the K-Means problems is based on a simple iterative scheme for finding a locally minimal solution is called K-Means algorithm. The k-means clustering algorithm is commonly used in computer vision as a form of image segmentation. The results of the segmentation are used to aid border detection and object recognition. In this context, the standard Euclidean distance is usually insufficient in forming the clusters. Instead, a weighted distance measure utilizing pixel coordinates, RGB pixel color and/or intensity, and image texture is commonly used. The k-means algorithm assigns each point to the cluster whose center (also called centroid) is nearest. The center is the average of all the points in the cluster that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster.

A. Methods.

1. Description of the web application:

In this study, we evaluated an Internet-based application for supporting self-care of patients with diabetes type 2, called the Diabetes coach. The purpose of the application was to persuade patients to have a more active role in their own care by facilitating online education and monitoring; thus for self-managing their disease. Second aim was to make health communication between patients and their career easier and more time-efficient by facilitating secure contact. The application was developed by Medicinfo in 2007 in close collaboration with GPs, nurses, patients, functional classification from an information technology perspective and classification in terms of the organization of health care delivery.

IV - Research

A. Patient characteristics.

Table 1 presents an overview of the patient characteristics. Participating patients (N=50) were predominantly male (n=37, 74%) with a mean age of 62 years (SD=8.5). The youngest participant was 43 years, the oldest 80 years. The majority of the participants had a Dutch origin (n=40, 93%), were medium (n=22, 51.2%) to highly educated (n=16, 37.2%), married (n=34, 79.1%), retired (n=18, 52.9%), and living independently (n=43, 100%). Most had been diagnosed with diabetes for one to five years (n=23, 56.1%), had relatives with diabetes (n=27, 62.8%), and were treated with diet and tablets (n=41, 95.3%). Almost all participants (n=40, 95.2%) were (very) satisfied with their current diabetes treatment (measured before start usage of the web application). We also measured treatment satisfaction (n=42). Almost all patients were very satisfied (n=18, 42.9%) or satisfied (n=22, 52.3%) with their current treatment (not yet involved with Internet-based care). Two patients had no clear opinion (4.8%).

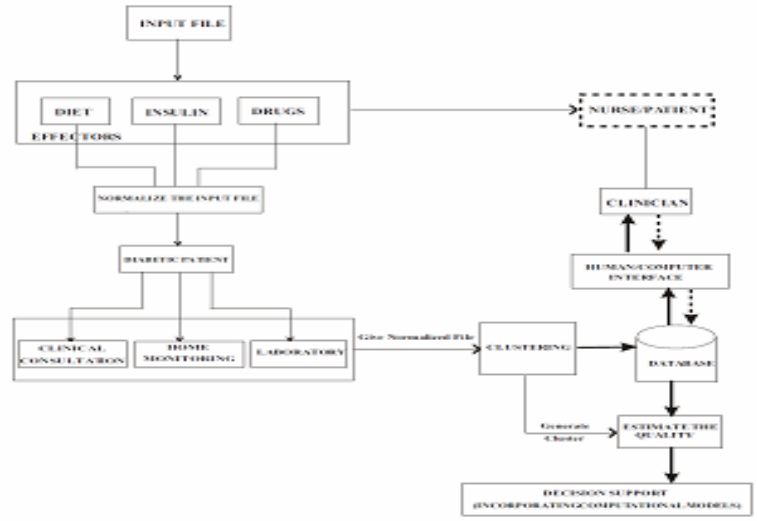

Fig. 2. The feedback loop of health care

Table: 1 Patient Characteristics

Characteristics		N	(%)
Gender	Male	37	-74
(N=50)	Female	13	-26
Education	Low	5	-12
(N=43)	Medium	22	-51
	High	16	-37
Origin	Dutch	40	-94
(N=43)	Other	3	-7
Marital Status	Not Married	2	-5.7
(N=40)	Married	34	-80
	Widower	2	-5.6
	Divorced	2	-5.7
Labor Activities	Paid Labor	10	-29
(N=40)	House Keeping	2	-5.9
	Unemployed	2	-5.9
	Incapacitated for Word	2	-5.9
	Retired	18	-53
Living Situations	Independently	43	-100
(N=43)	With Family/Friends	0	0
	In a Rest/Nursing Home	0	0

Table: 2 Patient Characteristics:

Characteristics		N	(%)
Diabetes Duration (N=41)	0-1 Year	1	-2.4
	1-3 Years	11	-27
	3-5 Years	12	-29
	5-7 Years	4	-9.8
	7-9 Years	6	-15
	9-11 Years	7	-17
Diabetes in family?	Yes	27	-63
	No	11	-26
	I Don't know	5	-12
Diabetes Treatment (N=43)	No Treatment	2	-4.7
	Diet	4	-9.3
	Diet and Tablets	37	-86
	Diet, Tablets and Insulin	0	0
	Other	0	0
Health Status (N=43)	Excellent	0	0
	Very Good	6	-14
	Good	25	-58
	Fair	12	-28
	Poor	0	0

1. A systems view of decision support:

Good systems design should always begin with a rigorous requirements analysis. There is the need to determine precisely what is the purpose of the intended decision support system, what the full range of potential user is and what are their real, rather than perceived, needs. Only once this has been satisfactorily completed, can the process of designing an appropriate decision support system begin. Fig. 2 focuses on the principal means by which the state of the patient is sensed together with the major categories of effector.

The opportunities for decision support can be categorized in a number of ways. These include the frequency: timescale of the clinical decision, ingredient. This systems perspective as described in this paper has developed and matured; so much so that it is reasonable to suggest that it will now become a standard framework to be adopted wherever there is the serious intent to design and produce decision support systems that meet the real needs of all their users and fulfill their purpose in an efficient, efficacious and effective manner. Further challenges remain, however, particularly in relation to the evaluation activity.

Although the case for a multi-dimensional perspective has been clearly demonstrated, of more is still needed in terms of developing tools with which to formally evaluate the worth of alternative configurations given that the systems approach in terms of stakeholders, criteria and dimensionality has already been accepted. Such development will undoubtedly take place so as to complete the array of systems ingredients necessary to ensure that the best possible decision support is made available to support the clinical management of diabetes mellitus.

V - Summary

This paper has sought to provide a systems perspective on decision support systems. The focus has been upon the delivery of care in the management of diabetes mellitus, though the concepts and approaches are relevant across the spectrum of chronic disease and beyond. Interest in the development of a decision support facility has a history of more than 20 years, and over this period a greater degree of maturity in the provision has become evident; as also the range approaches which now include stand-alone systems, integrated configurations and telemedicine systems. In parallel there has emerged the realization that clinical and technical expertise on their own are not sufficient to guarantee success; systems thinking and its translation into effective systems methodology constitute an additional, virtually essential

Internet into clinical practice as an information-sharing and communication medium has brought about many opportunities for providing immediate, transactional feedback on lifestyle modification, next to regular ways health care delivery. As such, Internet based applications can be used as a powerful medium for promoting healthy behavior, for self-care and also for improved care coordination. A primary focus of self-care applications for chronic illness is the encouragement of an individual's behavior change necessitating knowledge sharing, education, and understanding of the condition. But to be usable in empowering patients' self-awareness, Internet-based applications should be designed to allow Individuals to tailor the application to their specific needs, because patients increasingly demand convenient access to a high level of personalized health care. As a result, interactive and responsive e-health applications with continuous self-monitoring, treatments, and information exchange are more and more being developed. In Section II present the related work with K-means clustering algorithm and Diabetes mellitus, A proposed k-means clustering algorithm and discussed in Section III, in Section IV methods and Patients characteristics are represented. Summary are presented in Section V, Finally Section VI gives the Conclusion.

A. Nature of the problem—a dynamic systems Perspective.

The dynamic nature of managing diabetes is evident at more than one level. First, if one considers the healthy individual, an elevated level of blood glucose, for example as occurs following a meal, causes the pancreas to secrete insulin which, by means of its effects upon a range of physico-chemical processes, in turn causes the elevated blood glucose to be lowered. This is an example of a classic negative feedback loop. Insulin is the output variable of the pancreas acting as the controller. Blood glucose concentration is the output variable of the controlled system and is the variable which is sensed by the controller in this simple model (see Fig. 1). Original and back-translated English versions of the instruments were assessed by the research team and when there were semantic differences between two instruments, the Korean versions of instruments were revised and re-translated. After repeating the process, when the instruments were acceptable, readability was tested with five patients with diabetes mellitus, then final versions of the instruments were ready to use.

The comprehensive physical activities for the past days, including leisure time, indoor and outdoor work and transportation related activities were measured. Information about frequencies and duration of each activity were collected and scores summated to reflect the total physical activities conducted.

The revised Scale consisted of 12 items including self-care activities of diet, exercise, blood glucose testing, foot-care, and smoking. The measure asked the participant the number of days per week the participant had practiced self-care activities: '0' would indicate no performance at all, while '7' indicated a daily performance. Life style was measured by obesity evaluated from BMI (body mass index), status of regular exercise, amount of alcohol drinking, smoking status, and physical activity. BMI was calculated by weight and height measurements. The status of exercise, alcohol drinking, and smoking were measured by self-reported measures.

Self-efficacy on diabetic management behaviors was measured by SE-Type 2 (Bijl et al., 1999). The SE-Type 2 measured the degree of self-efficacy on maintaining diabetic diet, weight control, nutrition management, foot care, medical management for diabetic care, physical activity, Some of the most recommended methods of treatment include having frequent meals, monitoring your blood sugar levels and having foods that are low in simple and blood-glucose levels. Cronbach's Alpha of the scale in the study was found to be 0.94. The blood glucose levels were measured by random blood-glucose and HbA1c testing. Blood glucose was determined by a Hexokinase technique using the Hitachi 7600-110, 7170. HbA1C was determined by a High Pressure Liquid Chromatography (HPLC) technique using the Hitachi 7600- 110, 7170.

1. An historical perspective

The application of systemic methods and determining the characteristics. The estimation of the quality of a cluster helps us in identifying the algorithm that forms good quality clusters. Thus the characteristics of the diabetes data (available in form of clusters) can effectively be evaluated by applying the technique of ***Attribute Oriented Induction*** for the clusters generated by the identified algorithm (i.e., algorithm that generates good quality clusters).

Different algorithms are considered namely K-Means, Minimum Spanning Tree (MST) and Nearest Neighbor for generating the clusters and their quality is determined to identify the best algorithm that generates good quality clusters.

B. Study setting and participant.

Three primary care practices in the Netherlands started with the use of the web application in July 2007. A one year pilot was set up to evaluate the implementation, use and added value of the web application among its users. In total, 50

patients with diabetes type 2 and 6 nurses agreed to participate in the pilot project. Patients were recruited via a letter from their primary care practice or were asked by their career during an office visit.

Algorithm: MSTC ( )
Input : Protein sequence data set S
Output : optimal clusters

Let e1 be an edge in the EMST1 constructed from S
Let e2 be an edge in the EMST2 constructed from C
Let W_e be the weight of e1
Let σ be the standard deviation of the edge weights in EMST1
Let S_T be the set of disjoint subtrees of EMST1
Let n_c be the number of clusters

1. Construct an EMST1 from S
2. Compute the average weight of $\hat{W}$ of all the Edges from EMST1
3. Compute standard deviation σ of the edges from EMST1
4. $S_T = ø$; $n_c = 1$; $C = ø$;
5. Repeat
6. For each e1 $\in$ EMST1
7. If ($W_e > \hat{W} + \sigma$) or (current longest edge e1)
8. Remove e1 from EMST1
9. $S_T = S_T \cup \{ T' \}$ // T' is new disjoint subtree
10. $n_c = n_c + 1$
11. Compute the center C_i of T_i using eccentricity of points
12. $C = U_{Ti} \in S_T \{C_i\}$
13. Construct an EMST2 T from C
14. E_{min} = get-min-edge (T)
15. E_{max} = get-max-edge (T)
16. $CS = E_{min}/E_{max}$
17. Until CS < 0.8
18. Return optimal clusters

VI - Conclusion And Discussion

The web application is seen as a useful and worthwhile supplement to regular diabetes care. It proved to be a useful instrument for optimizing the self-care management of patients, because the application provided reliable and useful possibilities to learn about the disease and about disease control. Career gave feedback on measurements (alert values) and lifestyle issues, and made compliments on patients' healthy behavior. This continuous received feedback and support was greatly appreciated by patients; they felt more controlled and it motivated them to take a more active role in self-managing their disease. Successful diabetes management support systems should integrate several functions to meet the needs of patients and caregivers and require functionalities that have strong links to patients' existing clinical care.

The hierarchical k-means clustering algorithm, under investigation in this study will be redesigned based on the reported findings. In order to ascertain that the web application will be extended to more patients in other primary care practices and to other patient groups (insulin users, children, etc.), a supportive health policy environment and appropriate financing are necessary to guarantee continuity after the pilot period.

For Internet-based care applications, it is important to give adequate training and to explore which technology is best suited for whom and what changes are necessary to reach non-users or drop-outs [25]. Innovations in health care diffuse more rapidly when technology is used that is simple to use and has applicable components for interactivity and feedback in order to foresee in patients' need for a continuous healing relationship.

REFERENCES

[1] T. Bodenheimer, K. Lorig, H. Holman, and K. Grumbach, "Patient self-management of chronic disease in primary care," JAMA 2002;288(19):2469-2475.

[2] H.G. McKay, R.E. Glasgow, E.G. Feil, S.M. Boles, and M. Barrera Jr., "Internet-based diabetes self- management and support: initial outcomes from the diabetes network project," Rehabil Psychol 2002;47(1):31-48.

[3] F. Verhoeven, L. van Gemert-Pijnen, K. Dijkstra, N. Nijland, E. Seydel, and M. Steehouder, "The contribution of tele consultation and videoconferencing to diabetes care: a systematic literature review," J Med Internet Res 2007;9(5):e37.

[4] R.E. Glasgow, S.M. Boles, H.G. McKay, E.G. Feil, and M.Barrera Jr., "The D-Net diabetes self-management program: long-term implementation, outcomes, and generalization results," Prev Med 2003;36(4):410-419.

[5] H.I. Goldberg, J.D. Ralston, I.B. Hirsch, J.I. Hoath, and K.I.Ahmed, "Using an Internet co management module to improve the quality of chronic disease care," Jt Comm J Qual Saf 2003;29(9):443-451.

[6] M.M. Cassell, C. Jackson, and B. Cheuvront, "Health communication on the internet: an effective channel for health behavior change?" Health Commun 1998;3(1):71-79.

[7] D.J. Wantland, C.J. Portillo, W.L. Holzemer, R. Slaughter, and E.M. McGhee, "The effectiveness of web-based vs. non-webbased interventions: a meta-analysis of behavioral change outcomes," J Med Internet Res 2004;6(4):e40.

[8] R.E. Izquierdo, P.E. Knudson, S. Meyer, J. Kearns, R. Ploutz-Snyder, and R.S. Weinstock, "A comparison of diabetes education administered through telemedicine versus in person," Diabetes Care 2003;26(4):1002-1007.

[9] Asha Gowda Karegowda , M.A. Jayaram, A.S. Manjunath ,"Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients " ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012.

[10] 1Fahimuddin. Shaik 2Dr.Giri Prasad M.N.1Department of ECE, AITS, "An application of image segmentation and gradient filter methods in detection of atherosclerosis & exudative maculopathy in diabetic patients". National Journal on Electromic Sciences and Systems, Vol. 1, No.2, October 2010.

[11] Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, Senior Member, IEEE, An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 7, JULY 2002.

