Advanced Database Technologies in a Diabetic Healthcare System

Wynne Hsu, Mong Li Lee, Beng Chin Ooi, Pranab Kumar Mohanty, Keng Lik Teo, Chenyi Xia

School of Computing
National University of Singapore
{whsu, leeml, ooibc, dcspkm, teokengl, iscp0167}@comp.nus.edu.sg

Abstract

With the increased emphasis on healthcare worldwide, the issue of being able to efficiently and effectively manage large amount of patient information in diverse medium becomes critical. In this work, we will demonstrate how advanced database technologies are used in RETINA, an integrated system for the screening and management of diabetic patients. RETINA captures the profile and retinal images of diabetic patients and automatically processes the retina fundus images to extract interesting features. Given the wealth of information acquired, we employ novel techniques to determine the risk profile of patients for better patient care management and to target significant sub-populations for more detailed studies. The results of such studies can be used to introduce effective preventive measures for the targeted sub-populations.

1. Introduction

Diabetic retinopathy is a major cause of blindness in economically active adults worldwide. Regular screening of diabetic patients to detect the early stages of diabetic retinopathy so that appropriate and timely treatment can be given is the key to reducing the incidence of impaired vision and blindness from this condition. Current methods of detection and assessment of diabetic retinopathy are manual, expensive, and require highly trained personnel to read large numbers of fundus images, many of which will be normal. The efficiency of the screening programme can be improved by automating the initial task of analyzing the huge amount of retinal fundus images for symptoms of diabetic retinopathy.

Proceedings of the 28th VLDB Conference, Hong Kong, China, 2002

We have developed RETINA, a total-care system for mass screening of diabetic patients. The number of diabetic patients in the United States alone is close to 16 millions. In the UK, over 1.4 million people have diabetes and another million probably have the condition but do not know it. In Singapore, 10 percent of the population is diabetic. As a result, it is critical for such a system to handle large numbers of patient records and retina fundus images. In this work, we will demonstrate how advanced database technologies are being used in a seamless integration from capturing a patient's medical history and retina images, to the automatic analysis of images, to the risk profiling and sub-population studies of the patient database. Many state-of-the-art techniques have been employed in the system to make it user-friendly, efficient and scalable. In particular, we highlight the two unique features in the system:

1) Risk profiling using K-trees.

It is often necessary to know the risk profile of each patient in order to recommend a most effective care strategy for the patient. One way to determining the risk profile of a patient is to model it as a classification problem whereby past history will provide some useful patterns that capture the characteristics of patients in different risk groups. If a patient matches the characteristics of a particular risk group, his/her risk profile is updated to be the profile of the matched risk group. There are a number of problems with this approach. First, in the diabetic patient databases, the ratio of normal to high risk group is extremely biased. As a result, the classification patterns generated are unable to capture the profile of the high risk group accurately. Second, the patients are artificially being placed into different groups denoting different risks categories. This introduces unnecessary inaccuracies in the process. Third, the classification process needs to be re-run each time some patients changes their risk groups due to changes in their disease status which can occur very frequently.

To overcome these problems, we model the risk profiling problem as a new class of queries called the Reverse k Nearest Neighbors (RkNN). A novel indexing scheme, the K-tree, has been implemented to support RkNN queries efficiently. Details can be found in Section 3.

2) Sub-population targeting using CHive.

To ensure the effectiveness of total-care for the diabetic patients, it is important to develop good techniques that can quickly identify significant subgroups or clusters in the patient data. These subgroups or clusters may reveal interesting facts that allow preventive measures to be adopted, thereby improving the care of the diabetic patients on the whole. We term this problem as sub-population targeting. To support sub-population targeting, it is essential to have a good clustering technique that can quickly identify the natural groupings or clusters in a dynamic environment. We employ CHive, a novel clustering technique that we have developed for large volatile databases. CHive is able to support incremental clustering efficiently, and provide multi-resolution views of the data set. We have successfully identified significant sub-populations for further studies using CHive. Details are discussed in Section 4.

2. Overview of RETINA

Figure 1 gives the overall architecture and modules in the RETINA system. In this section, we will briefly discuss some of the capabilities in the system:

(a) Retina Electronic Medical Record (Retina-EMR)

The RETINA system has an underlying retina electronic medical record (retina-EMR) that provides for the documentation of patient demographic data, medical history, stage of retinal disease, treatment administered, patient care management, and serial images of the retina. The backend is built using SQL Server 2000 and the front end is served by Active Server Pages. Figure 2 shows a screen dump of a patient's Photography Report.

(b) Retinal Image Analysis

Part of the information in the retina electronic medical record comes from analysing the serial retina images of the patient. Novel image processing techniques [1, 2, 3] are employed to accurately identify abnormal retinal images based on a set of clinical criteria drawn up by a retinal specialist. The condition of abnormal images is quantified to allow various measurements to be made of the deterioration in the condition over time. The list of features detected includes optic disc/cup ratio, exudates, haemorrhages and vessel tortuosity. For example, an important manifestation of diabetic retinopathy is the development of hard exudates in the retina. We track the size, shape, and location of the hard exudates to allow automatic charting of the disease progression and to identify the risks associated with different types of progression patterns. The extracted features are inserted into the patient database. Figure 3 shows a screen dump for Retinal Image Analysis. All these information captured are subsequently used to determine a patient's risk factors for diabetic retinopathy, and subpopulation targeting. The details are elaborated in the next two sections.

3. Risk Profiling using K-trees

Traditional approaches to determine the risk profile of a patient are typically manual and tedious. In RETINA, we employ a novel risk profiling technique that allows customized risk assessment for each patient. This is achieved through first establishing the hypersphere of "influence" for each patient. The hypersphere of "influence" of a patient refers to the set of all patients who consider the given patient as one of their k^{th} closest neighbours in terms of the demographic, medical history, and retinal image data. This is also known as the Reverse k-Nearest Neighbours (RkNN) problem. We have developed a new access method called K-trees to process RkNN queries efficiently. The K-tree is unique in the following aspects:

(a) It does not prematerialize the RkNNs, and hence the update cost is low.
(b) It can give partial results quickly at the pruning stage, which alleviate the waiting time of a user in a real-time system.

Using K-Trees, we can efficiently query for the hypersphere of influence of any given patient. The risk profile of the given patient is then computed as the ratio of the number of patients in the hypersphere of influence who develop complications to the total number of patients in the hypersphere of influence. Initial testing shows that using RkNN queries to perform risk assessment does accurately reflect the actual risk profile of individual patients.

4. Sub-population Targeting using CHive

Given the potential wealth of clinical data and retinal images captured in the Retina-EMR, the next natural step is to discover interesting patterns and correlate them to disease outcomes. A typical patient record consists of hundreds of attributes. Making sense of patterns in such high dimensional space is almost impossible. To complicate matters, the patient database is highly volatile. New patients are being seen constantly and old patients may be referred out.

In order to solve this problem, we have developed CHive to efficiently obtain natural clusters in high dimensional yet volatile environment. CHive employs the concept of hyper-cube structure and density-based criteria to pick a set of representative points from the dataset. These representative points are used to form a Hive-graph that reveals the underlying cluster structure of a dataset. The Hive-graph localizes the effect of new data points on the current set of clusters. This reduces the number of clusters that require re-organization and eliminate the need to re-cluster from scratch. Furthermore, the Hive-graph can be manipulated to give a hierarchical view of the clusters, thus supporting multi-resolution clustering. Figure 4 shows a screen dump of the user-interface of CHive.

Each cluster identified by CHive corresponds to a sub-population in the diabetic population. By focusing on the significant clusters, we are able to target the significant sub-population for more detailed studies. The results of these studies may lead to new preventive or care management measures that can be adopted to improve the overall health care of the diabetic patients.

5. Conclusion

In this demo, we will show a practical system that employs advanced database technologies to achieve seamless integration, from capturing and indexing of the patient medical history and retina images, to the automatic computer analysis on the retinal fundus images of each patient, to determining the risk profile and sub-population targeting of the patient database, thereby providing total-care to all diabetic patients. This system represents the future of health care systems, where the aim is not just managing patient data, but also to integrate and analyze the data so that it can deliver total care to the patients.

Acknowledgement

This project is funded by the NSTB-NUS research grant. We would like to thank our collaborators, Dr Kah Guan Au-Eong (Department of Opthalmology, Tan Tock Seng Hospital), Dr Shanta Emmanuel and Dr Jonathan Phang (National Healthcare Group Polyclinics), for their valuable feedback and assistance in collecting the patient data. We would also like to thank the Canon support staff for agreeing to sponsor a fundus camera for the live demo at the VLDB conference.

References

[1] Huan Wang, Wynne Hsu, Kheng Guan Goh and Mong Li Lee. An Effective Approach to Detect Lesions in Color Retinal Images, in IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, USA, June 2000.

[2] Wynne Hsu, Mong Li Lee and Kheng Guan Goh. Image Mining in IRIS: Integrated Retinal Information System (Demo), in Proceedings of ACM SIGMOD International Conference on the Management of Data, Dallas, Texas, May 2000.

[3] Wynne Hsu, P. M. D. S. Pallawala, M. L. Lee and K. G. Au-Eong, "The role of domain knowledge in the detection of retinal hard exudates," IEEE Computer Vision and Pattern Recognition, Dec 2001, Hawaii.

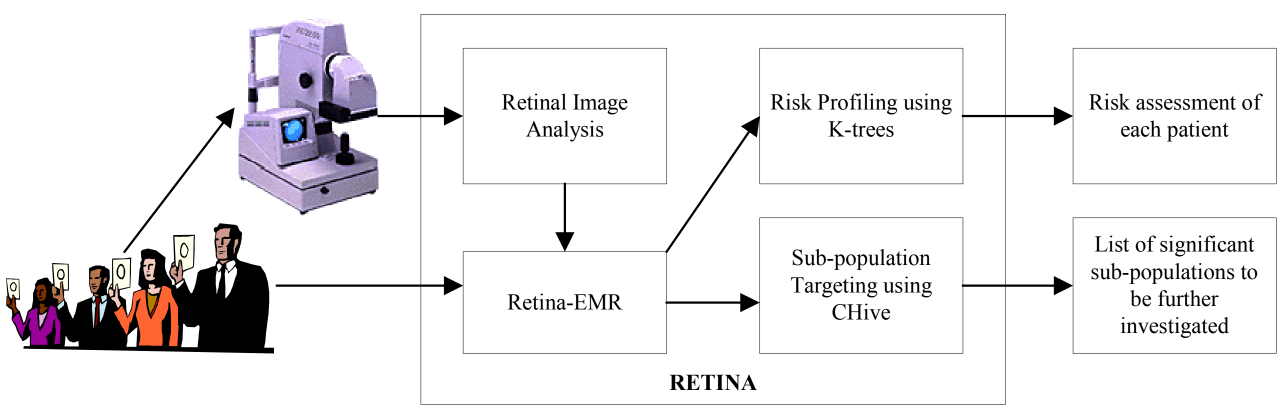

Figure 1. Overview of RETINA

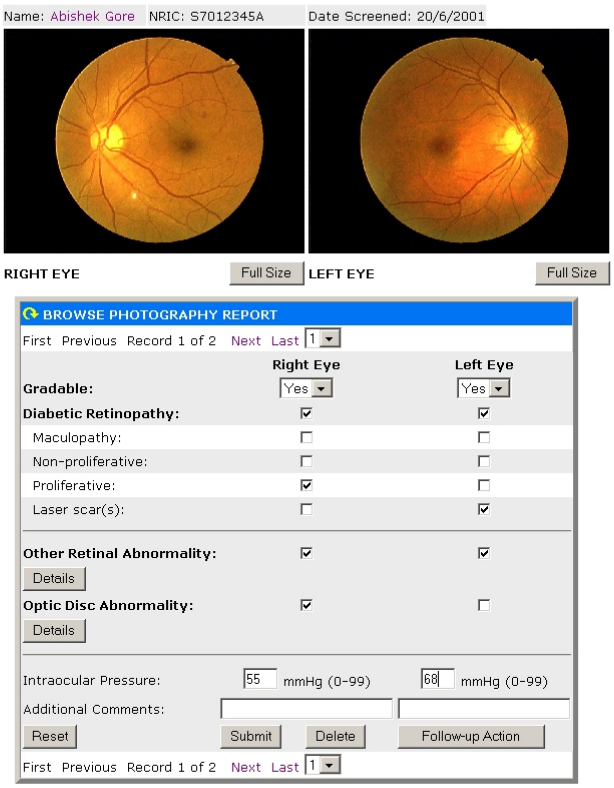

Figure 2. Screen Dump of Photography History.

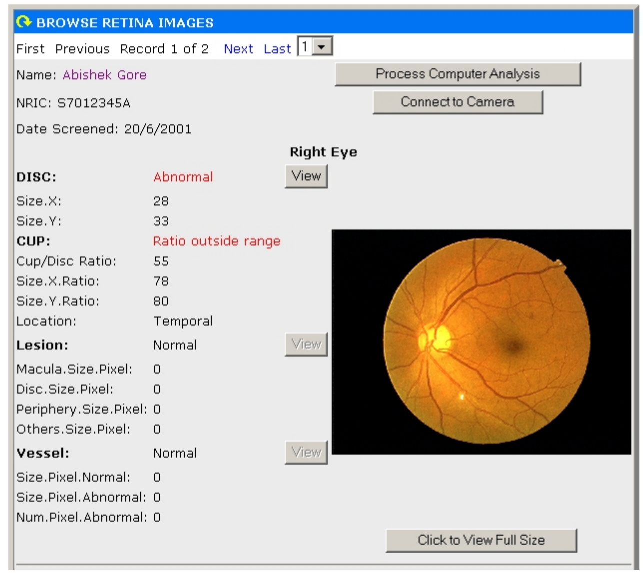

Figure 3. Screen Dump of Retinal Image Analysis.

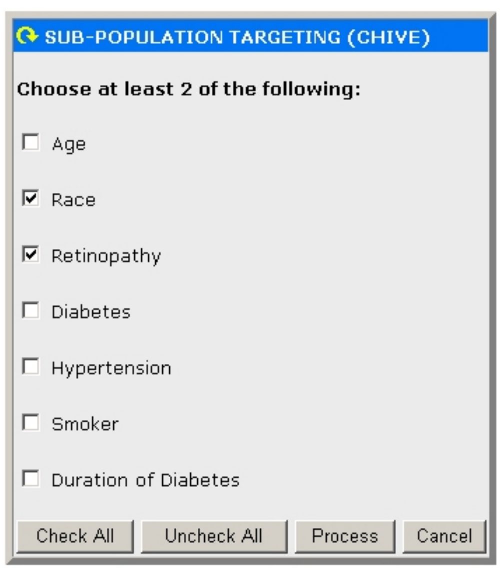

Figure 4. Screen Dump of Sub-Population Targeting (CHive) User-Interface.