Computer Methods for Sickle Cell Disease Research

Isaac Olanrewaju[1], Olumide Owolabi[2], Obiageli Nnodu[3]
[1,2]Computer Centre, [3]Department of Haematology
University of Abuja

Abstract

The use of computers for the storage, handling and analysis of medical research data has been well established. In recent times, however, there are some new approaches that have emerged that go beyond these basic computer data handling approaches. These methods range from the use of intelligent database systems to data mining and computer modelling approaches. In this paper, we review some of these methods and also outline some possibilities that are open to sickle cell disease researchers in Nigeria as concerns the application of computer-based methods.

1. Introduction

While the computer had been applied to various uses in medicine since its advent (DeTore, 1988), the launch of MEDLINE and MYCIN in the 1970s heralded a new era in medical applications of computers. Launched in 1971, MEDLINE is recognized as the first major online source of bibliographic and abstract coverage of the medical literature. It is the online version of MEDLARS, which had been in existence since 1964. It has since been followed by EMBASE, MedlinePlus and other such solutions. MYCIN was the first major intelligent computer application in the medical field. Drawing on earlier work in the area of Artificial Intelligence, MYCIN is an expert system designed to advise physicians on the selection of appropriate treatments for patients with bacteremia or meningitis. Others soon followed, such as: the Present Illness Program (PIP), a system that gathered data and generated hypotheses about disease processes in patients with renal disease; INTERNIST-1, a large system to assist with diagnosing complex problems in general internal medicine, and CASNET, an ophthalmology advisor designed to assess disease states and to recommend management for patients with glaucoma (Shortliffe, 1986). In recent times, the field of medical decision support systems using expert systems and other computer technologies are receiving much attention, making the field of medical informatics one of the best examples of interdisciplinary research.

2. Recent Advances in Computer Methods in Sickle Cell Disease Research

Data Mining

Data Mining involves the use of significant amounts of collected data to train a computer to learn hidden patterns in the data or make decisions about new cases. It has been described as "the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner" (Hand, Mannila & Smyth, 2001). Data mining methods include clustering, classification, neural networks and other machine learning algorithms. Data mining algorithms have been applied in the classification of sickle cell patient records, in some cases numbering up to a million records, to classify the cases and make decision on their treatment regimes (Solanki, 2014; Khalaf, Hussain, Keight & Tso, 2016; Solanki, 2016). An interesting case of the application of text and datamining in sickle cell disease research is the development of an information exploration system, Dragon Exploration System for Sickle Cell Disease (DESSCD), which aims to promote the easy exploration of SCD related data. The system processed 419,612 MEDLINE abstracts retrieved from a PubMed query using SCD-related keywords. The processed SCD-related data was then made available via the DESSCD web query interface that enables: information retrieval using specified concepts, keywords and phrases, and the generation of inferred association networks and hypotheses (Essack, Radovanovic & Bajic, 2013).

Computer Modelling.

A group at Brown University has been developing computer models to study interactions the sickling red blood cells lead to the blockages that cause the painful crises suffered by the patients. The goal is to be able to determine how to effectively apply drugs to minimize the effects of the crises (Lei and Karniadakis, 2013). A new method – the Mesoscopic Adaptive Resolution Scheme (MARS) – that overcomes the limitations of earlier methods by using some smart algorithms to handle the massive amounts of data involved in the study has just been published (Lu, Lei, Li and Karniadakis, 2017). High performance computing systems and novel modelling methodologies as these ones are likely to keep driving growth in this application area.

Medical Imaging and Abnormality Detection

Pattern recognition is another area of computing that found early application in the field of medicine. Computer programs are used to study images of different areas of the body in order to recognize the presence or otherwise of any anomalies in the part of the body being observed. Applications range from the detection of anomalies in blood cells to the classification of tumors (Benamrane, N., & al., 2005). The increasing availability of high-powered imaging systems, and large capacity data storage and processing systems is making this a fertile area of intersection between computing and medicine (Duncan & Ayache, 2000; Mehdy, Ng, Shair, Saleh & Gomes, 2017)

3. Research Directions for Computer-Based Methods in Sickle Cell in Nigeria.

Detailed Database of Sickle Cell Patients

Researchers at the Centre of Excellence for Sickle Cell Disease Research and Training, Abuja (CESRTA) have commenced a pilot community based screening program in Gwagwalada Area Council including an education intervention program. This we expect will yield large demographic, clinical, social, laboratory, and research data over the course of these and future studies. An electronic database is being designed and implemented for the acquisition, storage, longitudinal tracking, and analyses of such large data sets. The development of the database to support the SCD research will proceed by first developing a list of the demographic, social, diagnostic, clinical, laboratory, imaging, and research data elements ("fields") to be included in the Database. The database will be incrementally developed using an appropriate relational database system to support the number of tables that can sufficiently represent all the data being gathered. REDCap (Research Electronic Data Capture), a secure, web-based application designed to support data capture for research studies, will be deployed to capture the data. Data Tables will include: Basic demographics of the screened population, result of screening, contact numbers, exact location, names of schools and nearest primary health care centres. For those with sickle haemoglobin, either SS SC basic patient data (demographics), routine clinic visits, acute care visits, hospital admissions and laboratory studies will be collected. Each Table, when developed with relevant data elements suggested by the analysis, will be translated into user-interface forms and field-tested at the data collection sites. Changes will be made in the design of the Table based on the results of the field test. Eventually, the database will be deployed on the web, with appropriate security features. The local versions of the database will allow secure on-site clinical use as well as collection and transmission of research data to the research team members.

Apply data mining methods to data to generate inferences

The data to be collected in the database will be analyzed to document the prevalence of the disease, the pattern of complications, morbidity, mortality and possible management options. Analysis of such data will reveal patterns and relationships in the data; and provide new insights into disease distribution and viable and effective interventions. Novel data mining methods such as classification and neural networks could be applied to the data.

Searchable Database of SCD Studies in Nigeria

In the long run, a national database will to record data and results by SCD research teams across the country and accessible at one point is highly desirable. A system similar to DESSCD (Essack,

Radovanovic & Bajic, 2013) will yield rich information about SCD research in different parts of Nigeria and will allow for comparisons and correlation of results. This will give a high visibility to the work being done by researchers in Nigeria. Thus the larger medical community will be aware of the situation in Nigeria and will provide comparative data that have the potential for broadening knowledge about the disease and contributing to management protocols worldwide.

Visualizations from Data

Visualizations are a great way to view data and see information dispersal over several regions or areas. The data collected as described above could be used to make visualizations to view at a glance the location of SCD researches and the prevalence across the country. Relationships between different factors could also be explored by the visualizations. The usefulness of this approach to the exploration of data has resulted in the creation of different tools for it. This will also form an interesting area of work for SCD researchers.

Database of SCD Researchers in Nigeria

The prosecution of this work will be greatly helped by the establishment of a database of SCD researchers in Nigeria. Profiles of researchers with their past and current projects and results should be published for effective collaborations among the researchers. Publication of the available equipment and methodologies adopted will also be of interest. We intend to circulate some Google forms for the collection of some of these data and, therefore, solicit for support in terms of members' emails and prompt responses. The database will be made easily searchable in order to benefit the entire community.

4. Conclusion

We hope that this short paper has given us some insights into works that have been done in the area of Medical Informatics and how, riding on some of these developments, we can leverage the power of information technology to work more productively.

References

Benamrane, N., & al. (2005). A Hybrid Fuzzy Neural Networks for the Detection of Tumors in Medical Images. *American Journal of Applied Sciences*, 2(4), 892-896.

DeTore, A.W. (1988). Medical informatics: an introduction to computer technology in medicine. *Am. J. Med.* 85(3):399-403

Duncan, J.S., & Ayache, N. (2000). *Medical Image Analysis: Progress over Two Decades and the Challenges Ahead*. IEFE Transactions On Pattern Analysis And Machine Intelligence, 22(1), 85-108.

Essack M., Radovanovic A. & Bajic V.B. (2013). Information exploration system for sickle cell disease and repurposing of hydroxyfasudil. , 8(6):e65190. doi:10.1371/journal.pone.0065190.

Hachaj, T. and Ogiela, M.R. (2013). Nowadays and future computer application in medicine. IT CoNvergence PRActice (INPRA), 1(1): 13-27.

Hand, D. Mannila, H. and Smyth, P. (2001). *Principles of Data Mining.* MIT Press, Cambridge. ISBN: 026208290x

Khalaf, M., Hussain, A.J., Keight, R. & Tso, P. (2016). Machine Learning approaches to the application of Disease Modifying Therapy for Sickle Cell using Classification Models. *Neurocomputing* 228. DOI:10.1016/j.neucom.2016.10.043

Lei, H., Karniadakis, G.E. (2013). Probing vasoocclusion phenomena in sickle cell anemia via mesoscopic simulations. *Proceedings of the National Academy of Sciences*; DOI: 10.1073/pnas.1221297110

Lu, L., Li, H., Bian,X., Li, X. and Karniadakis, G.E. (2017). Mesoscopic Adaptive Resolution Scheme toward Understanding of Interactions between Sickle Cell Fibers. *Biophysical Journal*, 2017; 113 (1): 48 DOI: 10.1016/j.bpj.2017.05.050

Mehdy, M.M., Ng, P.Y., Shair, E.F., Saleh, N.I. Gomes, C. (2017). Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. *Computational and Mathematical Methods in Medicine*. Article ID 2610628. https://doi.org/10.1155/2017/2610628

Shortliffe, E.M. (1986). Medical Expert Systems-Knowledge Tools for Physicians. *In Medical informatics [Special Issue]. West J. Med.* 145:830-839.

Solanki, A. V. (2014). Data Mining Techniques Using WEKA classification for Sickle Cell Disease *Int. J. of Computer Science and Information Technologies* 5 (4): 5857-5860

Solanki, A.V. (2016). Study and Analysis of Sickle Cell Disease of South Gujarat region with the use of Data Mining Classification Techniques. *Int. J. of Computer Technology & Applications* 7(5): 688-692.