Data Analysis Methodology for Sand Rat Model

Syed Raza, Marc A. Wilson, Leonard S. Cole, Shikha Gola, V. Michelle Chenault
Talladega College / Talladega College / Talladega College /Case Western Reserve University / Food and Drug Administration
sqraza@talladega.edu / Marc.Wilson@fda.hhs.gov / lcole@talladega.edu /Shikha.Gola@Gmail.com / v.chenault@fda.hhs.gov

Abstract

This project will analyze the methodology used to apply statistical functions and data log transformation by a proposed Sand Rat Data Analysis Module (software) SRDAM to assist the exportation problem encountered by the B1-9000AT Digital Autocorrelator program. A module or techniques were developed to export the multiple data files from BI-9000AT to the SRDAM software then into MS-EXCEL® software to overcome the problems encountered in BI-9000AT software. BI-9000AT does not have the ability to automatically perform statistical functions or the log transformation needed for data analysis or interpretation from data collected by a NASA developed fiber optic device. BI-9000AT is a digital, high speed, signal processor which is used as an auto or cross correlator for dynamic light scattering (DLS) measurements performed in the eye. The sand rat, *Psammomys obesus,* is unique in that it is a nutritionally induced animal model for Non-Insulin Dependent Diabetes Mellitus (NIDDM) or type 2 diabetes mellitus (T2DM). Complications seen in diabetics such as blindness due to cataracts can be studied using the sand rat as a model. DLS can be used to detect cataracts at the early stages. The BI-9000AT with the SRDAM software efficiently and accurately calculates several data files at one time. Thus, study data can be analyzed at the time of collection from the animals for the investigation of various aspects of ocular health- specifically lens opacity, as a possible early indcator of diabetes mellitus.

Background

The project described below is collaboration between personnel at multiple government agencies, faculty, and students from Talladega College and Case Western Reserve University to develop a software module that will allow data mining, statistical calculations and data analysis. This data will provide an unprecedented opportunity to use data mining techniques to facilitate non-invasive studies of changes in the eye in an animal model of diabetes that mimics the human syndrome. Diabetes mellitus affects approximately 6% of the population [1].Diabetes complications result in extensive end organ damage, such that one–third of all hospitilized patients are diabetic, 30% of the patients requiring kidney transplants have diabetic nepropathy, gangrene of the lower extremities is 100 times more common in diabetics, diabetic retinopathy is the fourth leading cause of legal blindness and 100 million eyes are legally blind from catartacts today [2], [3].The data collected from these investigations will add important information on the progression of ocular health during the development of diabetes and provide a new tool for early cataracts detection.

Introduction

Psammomys obesus or the fat sand rat is a wild desert rodent in the gerbil family. The occurrence of diabetes mellitus in the sand rat is a subject, which has aroused much interest since it was first reported [4]. The sand rat is a nutritionally induced animal model for Non-Insulin-Dependent Diabetes Mellitus (NIDDM), also called type 2 diabetes mellitus (T2DM).This animal is unique in that it develops mild to moderate obesity, hyperglycemia and complications from diabetes such as blindness, kidney disease, delayed wound healing and heart disease by dietary induction [4],[5],[6].

The sand rat dwells in the salty deserts of Algeria, Tunisia, Saudi Arabia, Israel, Sudan, Libya and Egypt. Their main diet in the wild is the low-energy, electrolyte rich salt bush plant, (*Atriplex halimus*) [6]. In the laboratory environment, the sand rat develops obesity, hyperglycemia and hyperinsulinema when fed a high caloric diet [5], [7], [8]. The development of cataracts in the sand rat is associated with the development of nutritionally induced T2DM (Figure 1). Various researchers have found that the diabetes syndrome in sand rats is relevant to pathogenesis of diabetes in certain populations including other minority groups in the United States (Arizona Pima Indiana and Micronesians) [9].

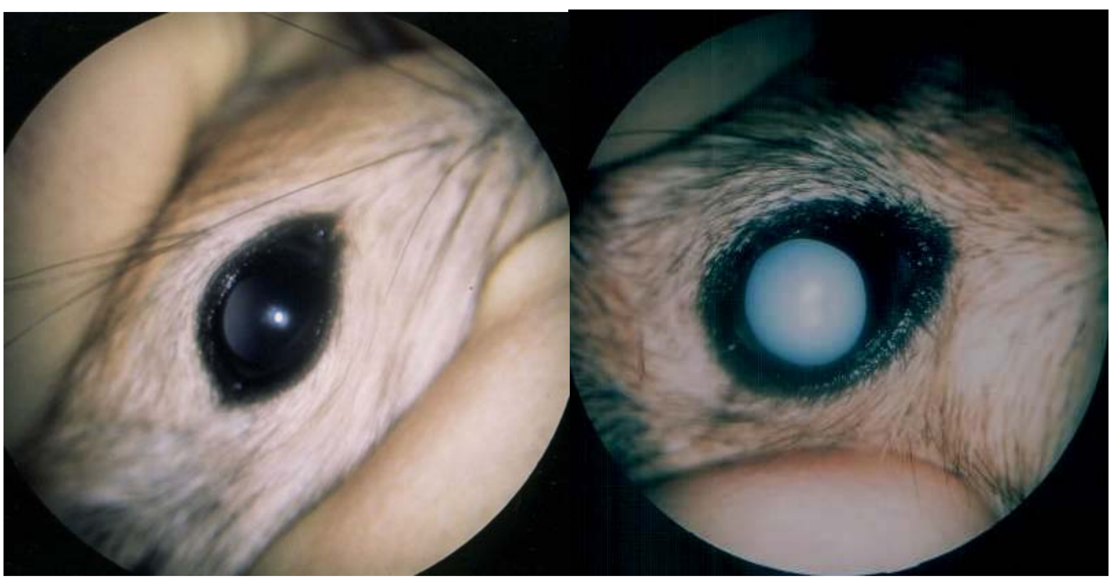

Figure 1: Photograph of eye fundus of *Psammomys obesus* (sand rat).Left eye is normal. Right eye shows a mature cataract.

The sand rat project focuses on the early detection of ocular damage induced by diabetes. Researchers (Chenault et al) [10] are able to view the pathologic or normal traits of the eye using Dynamic Light Scattering (DLS) fiber-optic probe developed by NASA scientists (Ansari, et al) [11], [12], [13] (Figures 2 and 3)**.** An optical fiber transmits a low- power laser beam into the eye [12], [13], [14]. Light scattered from within the eye back to the instrument is detected by a second optical fiber. Next, the data collected from the lens is stored, analyzed and processed into the summary file by the BI-9000AT Autocorrelator [11]. DLS data is depicted as protein particle size distribution. A time correlation function is used to calculate the particle size in the lens crystalline. A change in protein particle size (increase) in the eye indicates complications due to opacity [10], [14] (Figure 2).

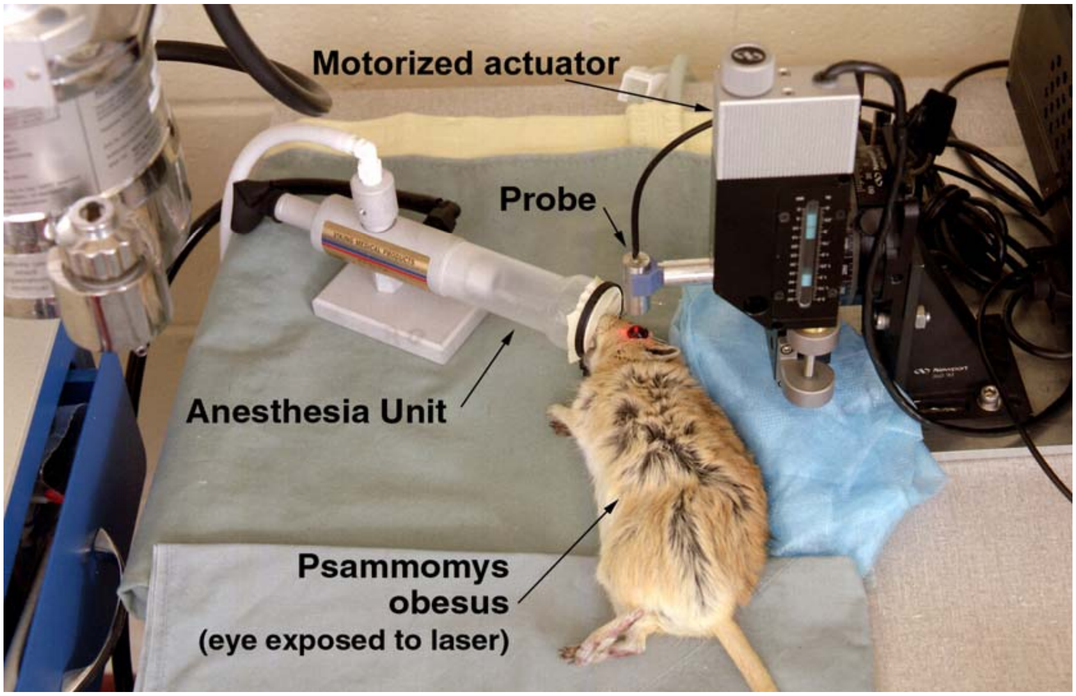

Figure 2: NASA's Dynamic Light Scattering (DLS) fiber-optic probe

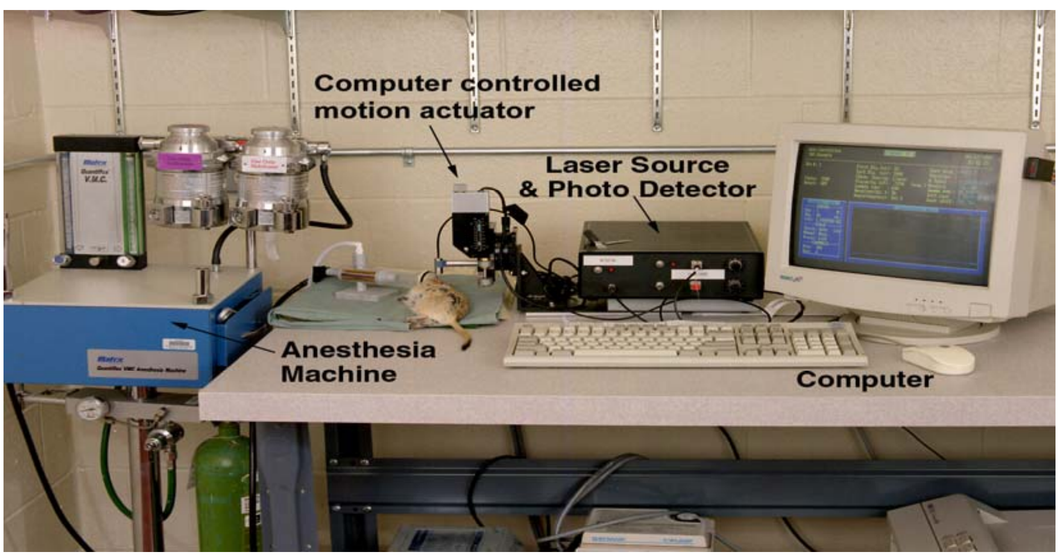

Figure 3: The laboratory set up for eye examination and data collection.

The animal is anesthesized for the procedure and all studies (performed by the principal investigator, V.M. Chenault) were conducted under an approved protocol at Uniformed Services

University of the Health Sciences in Bethesda, MD and in keeping with the regulatory guidelines for the humane care and use of animals (Figure 3).

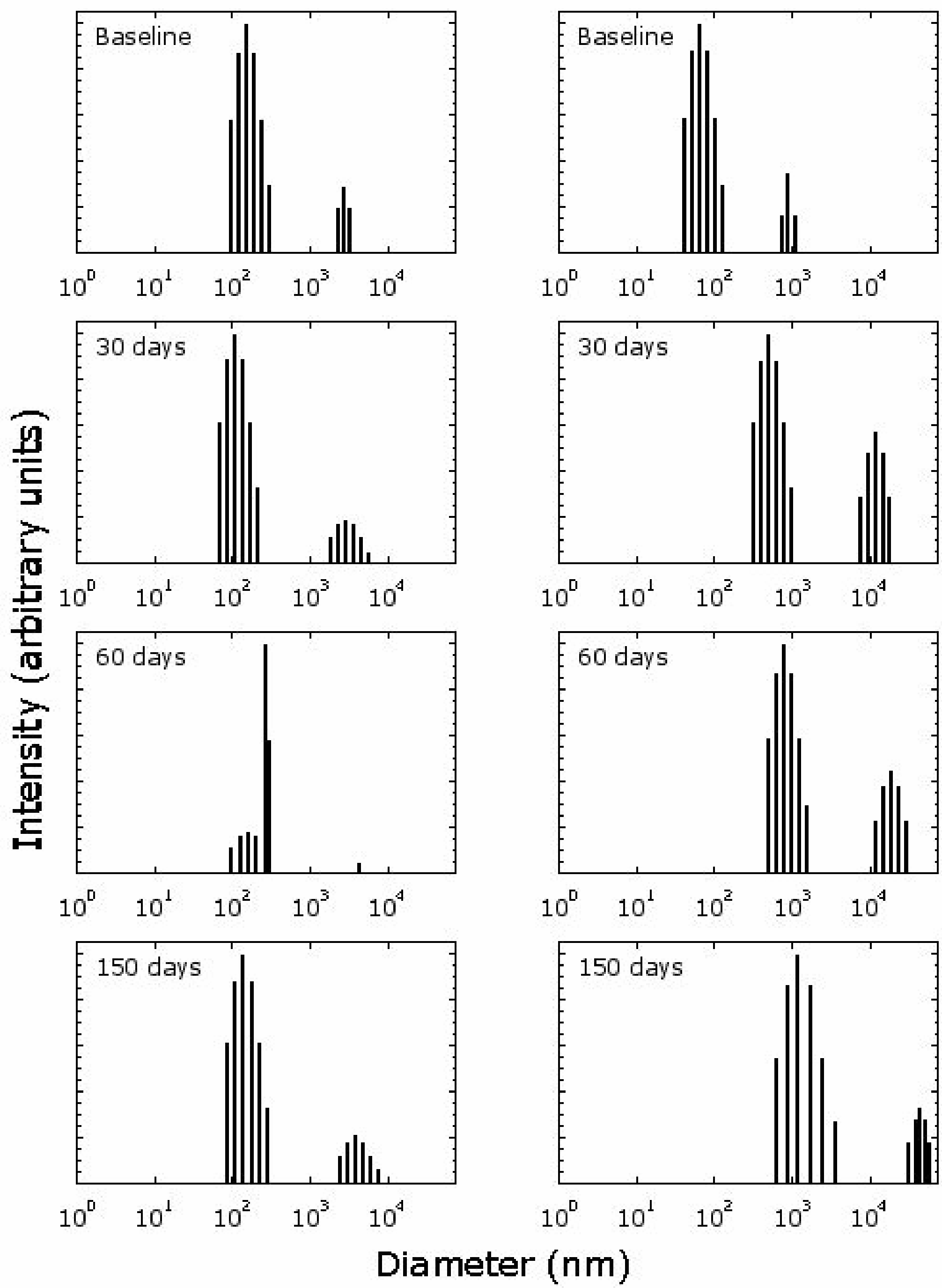

Figure 4: Sample particle size distribution data collected from analysis of the lens of the nucleus of the sand rat eye using DLS. Note there is a shift from low molecular weight to high molecular weight proteins [16].

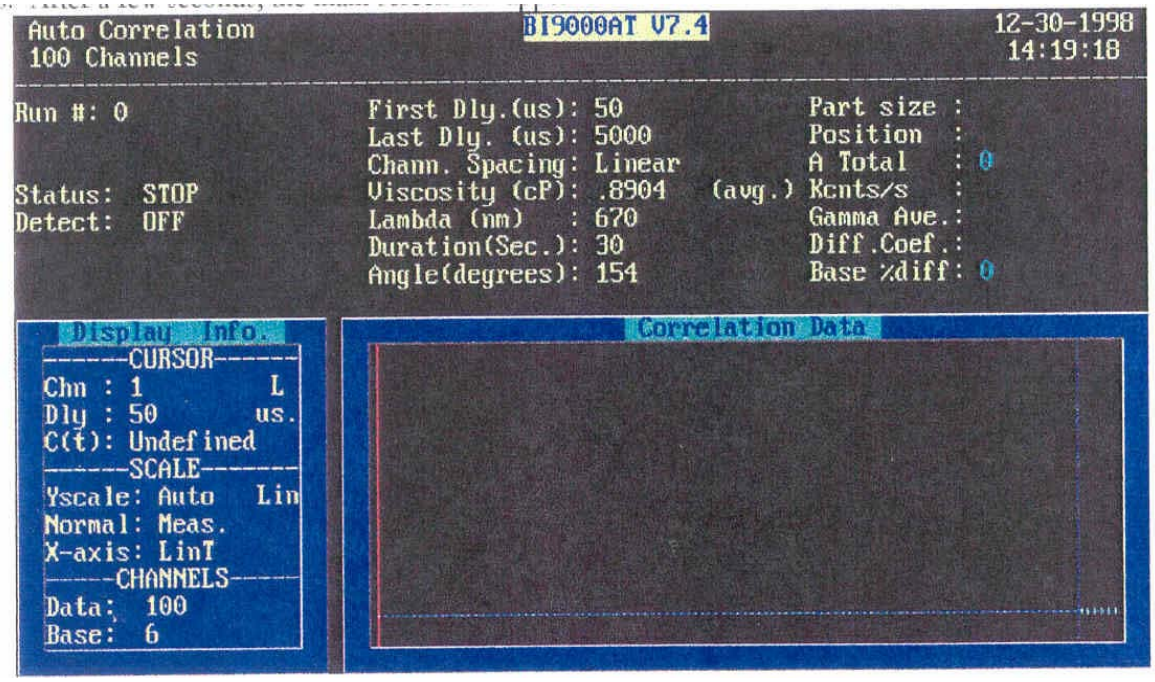

Figure 5: BI-9000AT Main Screen

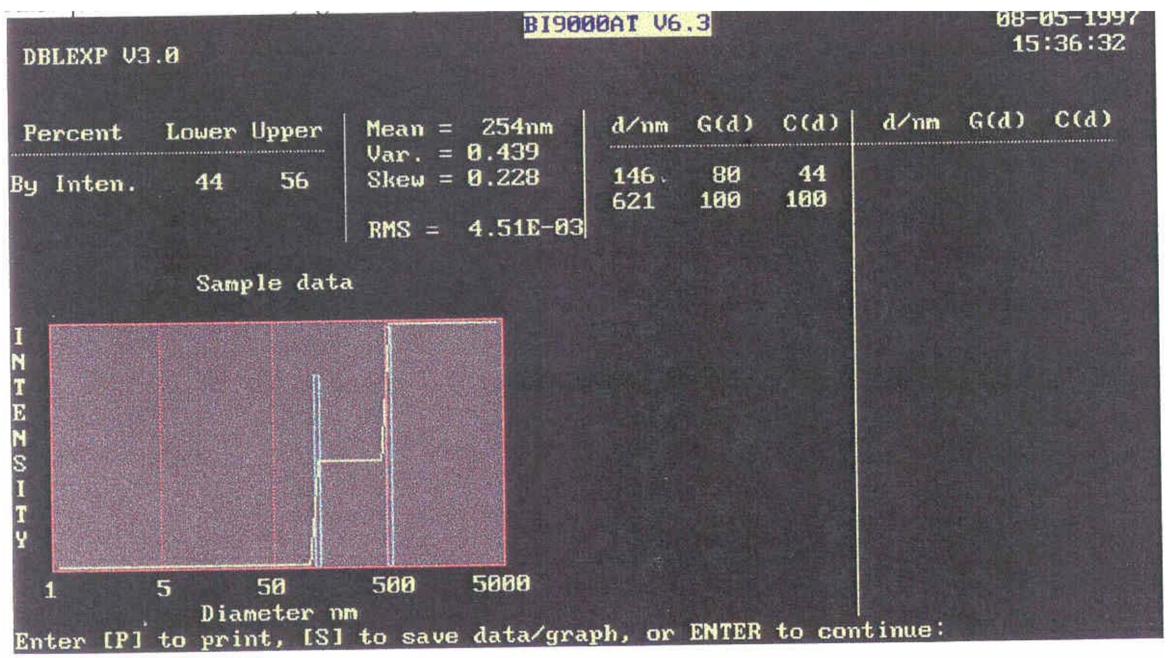

Figure 6: Double Exponent Result

The BI-9000AT is an entirely digital, high speed, signal processor that can be used for cross correlation calculations of dynamic light scattering (DLS) measurements. Its correlator control program is written in C++ language. Figure 5 shows a main screen for the BI-9000AT program. The box marked “Correlation Data” is where the correlation function is plotted. Figure 6 shows the screen for the Distribution Analysis through BI-9000AT program. Note that instructions are provided in this screen to print [P] data or save [S] data in a summary file.

The BI-9000AT consists of three important steps in order to generate the final results into particle size distribution data (Histograms as depicted in Figure 4). The three steps are:

- Step 1. Loading data files into the system
- Step 2. Creating Summary Files
- Step 3. Importing Summary File in MS-EXCEL®

The most critical steps in this software are between steps 2 and 3 in BI program. A fourth step is needed to perform log transformation into one valve that represents lens health or opacity. The importing of summary file into MS-EXCEL® is performed to create the BI-9000AT histograms.

The BI-9000AT lacks the ability to: 1) collect data summary file of DLS measurement; 2) transfer multiple data files into MS-EXCEL®; 3) pick individual d/nm columns from summary file and import to MS-EXCEL® [11].

The complexity arises between the steps previously mentioned and other needed features in order to produce efficient and accurate results in real time. The BI-9000AT program uses MS-EXCEL® to produce a distinct hardcopy of data and does not have the ability to perform logarithm transformation and statistical functions. The process for loading individual files into MS-EXCEL® is time consuming in order to produce results from DLS and for application of statistical functions. Multiple files have to be entered into MS-EXCEL®, individually, to produce results. Therefore, the BI-9000AT was not efficient because it was not sufficiently designed to produce results in an effective manner.

Methodology

The lead author (Raza) prioritized the problems and current complications with the BI program to ascertain the importance of each function's role in determining the results. Secondly, a feasibility analysis was performed to assess the technical feasibility, organizational feasibility and economic feasibility to meditate the risks of each methodology.

Data Analysis

The first methodology was to create C++ /C # module to accommodate the BI-9000AT program. This particular methodology proposes a sufficient time saving effort and reduces the risk. Ultimately, the methodology that the author chose is to create a module written in C# language by using SharpDevelop 2.0 (Integrated Development Environment -IDE for Visual Studio.Net®) from Microsoft Corporation. Statistical packages such as MS-EXCEL®, or Statistical Package for the Social Sciences (SPSS) [17] can be used to apply statistical functions. During the assessment of the technical feasibility a small risk appeared in respect to familiarity with the software. This risk is very low because the proposed module, based on Graphical User Interface (GUI), is somewhat effortless. The software module (Sand Rat Data Analysis Module- SRDAM) provides a graphical user interface, which will allow the user to easily navigate and perform desired steps by loading multiple summary files through one interface. By using this module, intangible benefits increase tremendously. SRDAM is versatile software that can be used for an organization's predictive analytical needs, statistics, data mining, and survey analysis or for decision optimization in the project. Figure 7 depicts the main screen of SRDAM software module, used to capture the data from BI-9000 AT program.

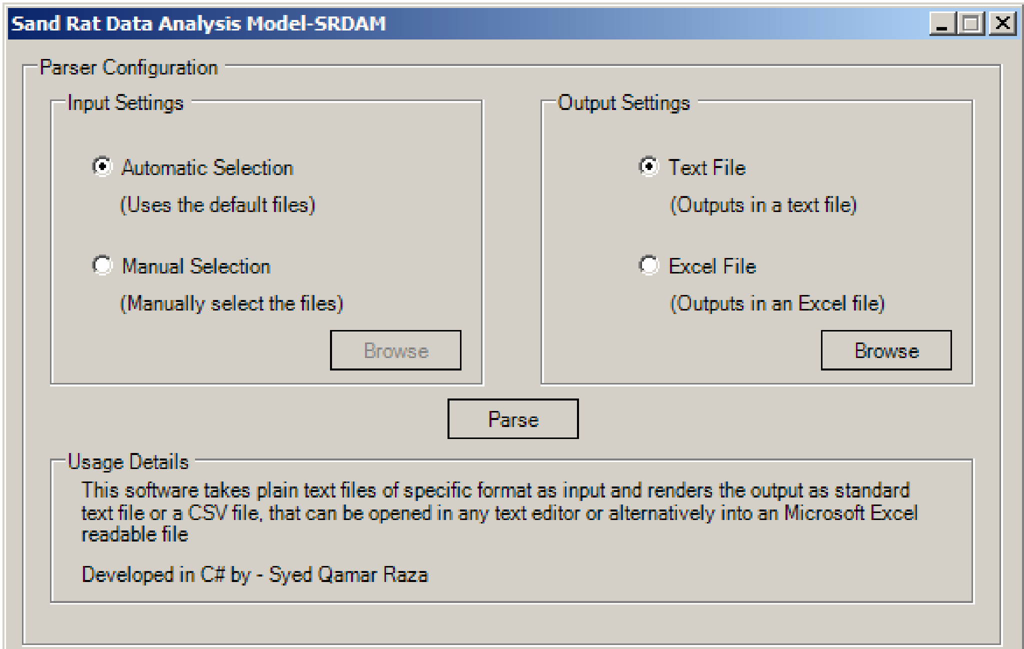

Figure 7: SRDAM software module main screen

Nomenclature of File System

The following list explains the nature of the files.

- DAT extension means "correlation data file".
- RAT extension means "count-rat data file".
- SUM or DUM means "summary file".

Results and Discussion

The methodology uses the following major steps to apply statistical functions to get the desired results in a timely manner.

Loading .SUM or .DUM files created by the BI-9000 AT program into the SRDAM software module was performed through the main GUI interface. That screen had an option to select default files. The default files were hard coded into the source code or selected manually by pressing the browse button. (see Figure 8, SRDAM Data Capture Screen- see notation on the left hand side).

After loading all summary files, the second step was to send the output into MS-EXCEL®. This could be performed by pressing the “Parse” button. At that point, the interface asked, for a file name. Once the file name was given, it picked all d/nm columns from the summary files and transported them into MS-EXCEL®. (Figure 8, Data Capture screen- see notation on the right hand side) Next, the MSEXCEL® file was opened and the Log Transformation steps, mean and Standard Deviation analyses were performed through MSEXCEL® formula.

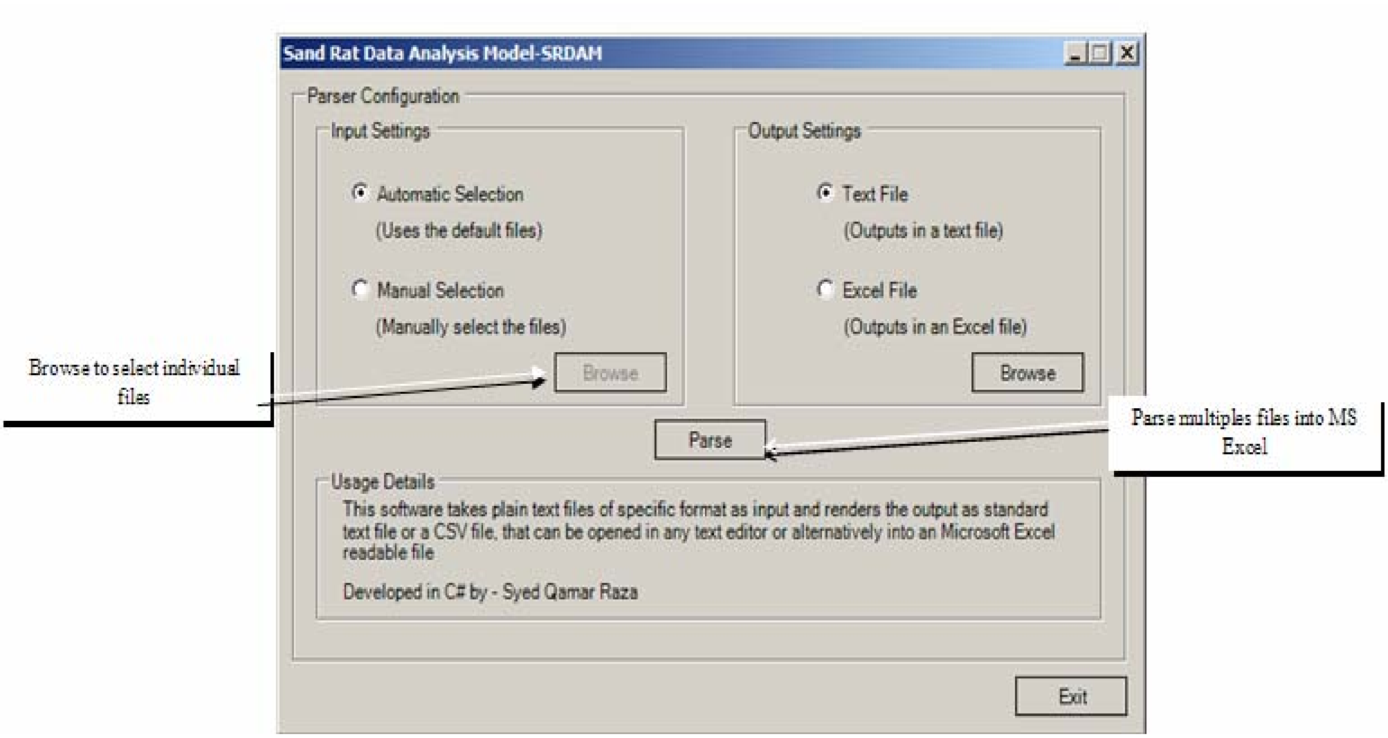

Figure 8: SRDAM Data Capture Screen

Figure 9, 10, 11 and 12 depicts sample screens for the steps required to produce the final results after assessment of the sand rat lens using the fiber optic device. Sample graphs are provided to depict the statistical functions (mean, standard deviation and the log transformation data point).

```
EXPSAM V3.0 (126N1)
Sand rat data
d/nm     G(d)      c(d)
  15                54          7
  18      76        18
  22        93      30
  27     100        44
  33         93        56
  41        76      66
  50        54      74
  62        33      78
  76      18        81
  94       8        82
 115       0        82
 149       0        82
 188       0        82
 256       6        82
 313      11        84
 384      17        86
 470      22           89
 575        24      92
 704      22        95
 862      17        98
1056      11        99
1292       6       100
1827       0       100
2295       0       100
```

Figure 9: Sample (raw) data for Summary File

	A	B	C	D	E	F	G	H
1	120SR1	120SR4	122SR4	126N1	126N3	126N4	126SR4	127N4
2	17	15	65	15	51	55	13	15
3	21	19	81	18	64	69	16	19
4	26	24	102	22	81	87	20	24
5	32	31	129	27	102	110	24	30
6	39	39	162	33	128	138	30	38
7	48	49	204	41	162	173	37	48
8	59	62	270	50	214	230	46	61
9	72	79	355	62	282	303	56	77
10	88	100	508	76	371	398	69	98
11	108	127	611	94	489	525	85	123
12	153	191	734	115	644	691	105	188
13	192	249	1070	149	848	911	130	244
14	241	323	1752	188	1225	1340	189	317
15	319	420	2204	256	1472	1670	260	450
16	391	662	2773	313	1769	2082	318	567
17	478	832	3489	384	2554	2596	389	716
18	585	1047	4390	470	3364	3236	477	904
19	717	1316	5523	575	4432	4759	584	1140
20	877	1655		704			715	1439
21	1074	2081		862			875	1816
22	1315	2617		1056			1071	2292
23	1610	3291		1292			1311	2893
24	2351	4428		1827			1846	4339
25	2952	5753		2295			2318	5638
26								
27	573.5417	1058.75	1356.778	455.1667	1014	1076.278	457.6667	978.1667
28	782.492	1545.404	1663.622	612.0002	1271.254	1325.986	620.9959	1474.874

Figure 10: Data capturing from Summary File

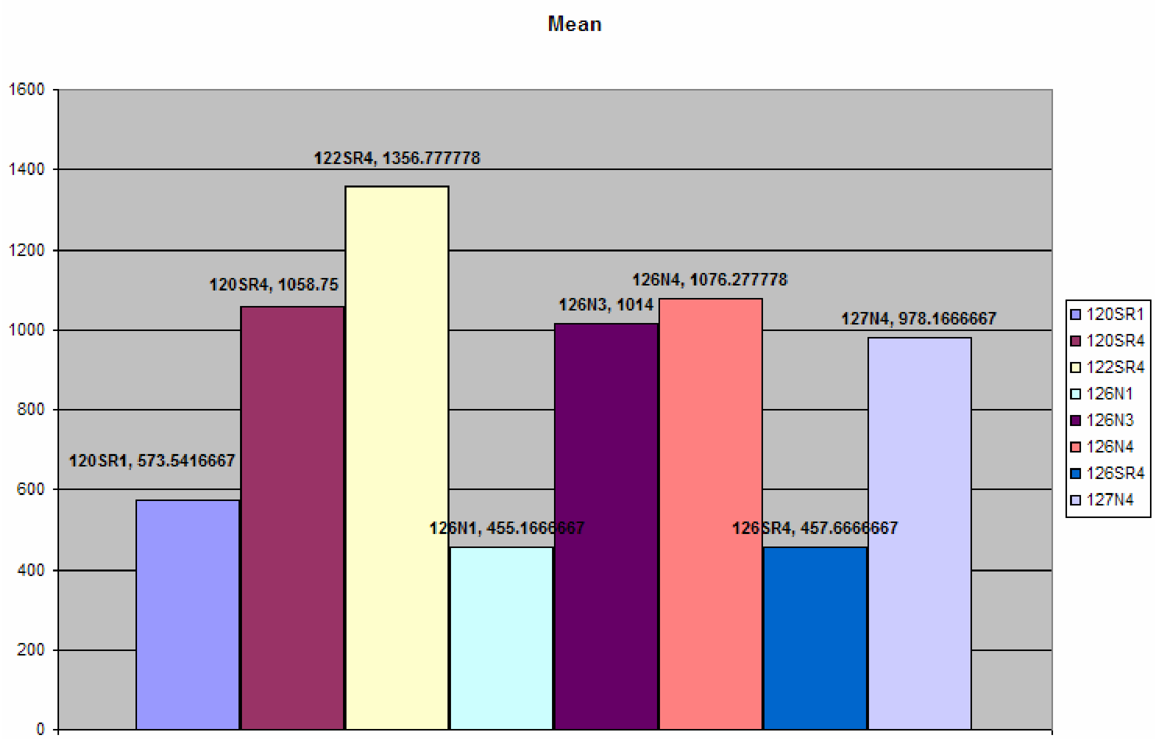

Figure 11: Mean DLS results representing lens particle size and opacity for 8 individual measurements of the lens nucleus from the data in Figure 10 (Row 27).

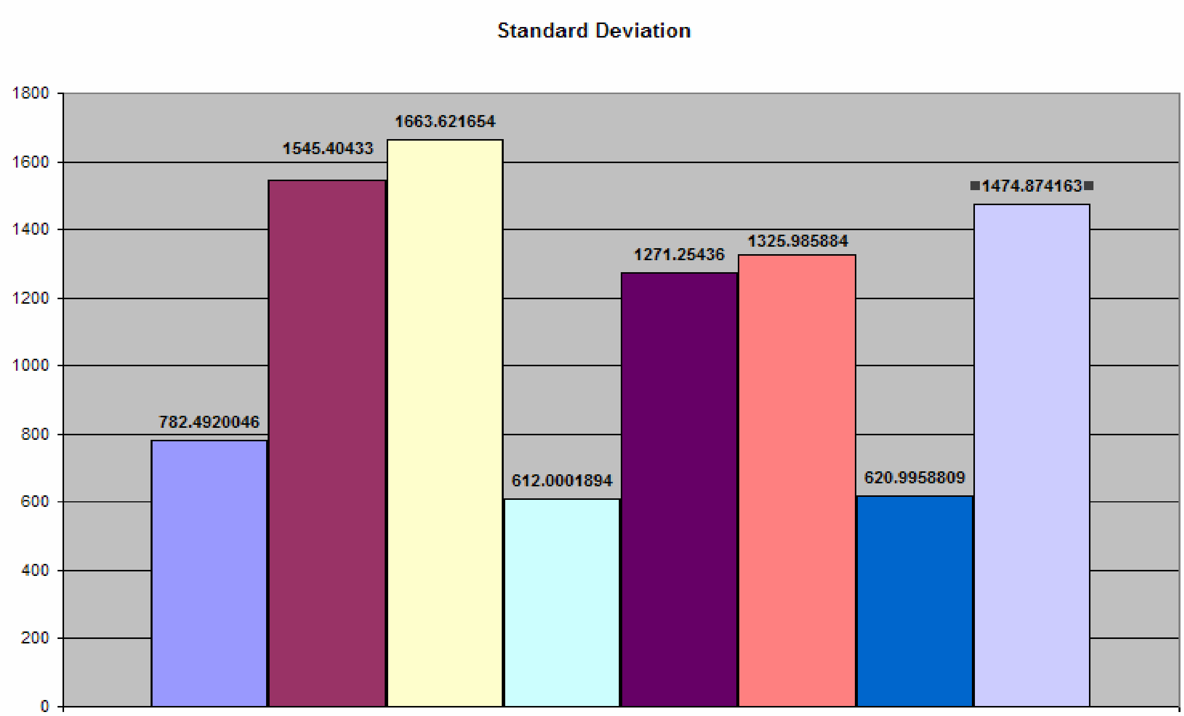

Figure 12: Standard Deviation calculations determined from the data depicted in Figure 10 (Row 28).

B	C	D	E	F	G	H	I	J	K
Files	**120SR1**	**120SR4**	**122SR4**	**126N1**	**126N3**	**126N4**	**126SR4**	**127N4**	
Mean	573.5417	1058.75	1356.778	455.1667	1014	1076.278	457.6667	978.1667	
Log Transformation	2.758565	3.0247934	3.1325088	2.6581705	3.006038	3.0319245	2.6605493	2.9904129	

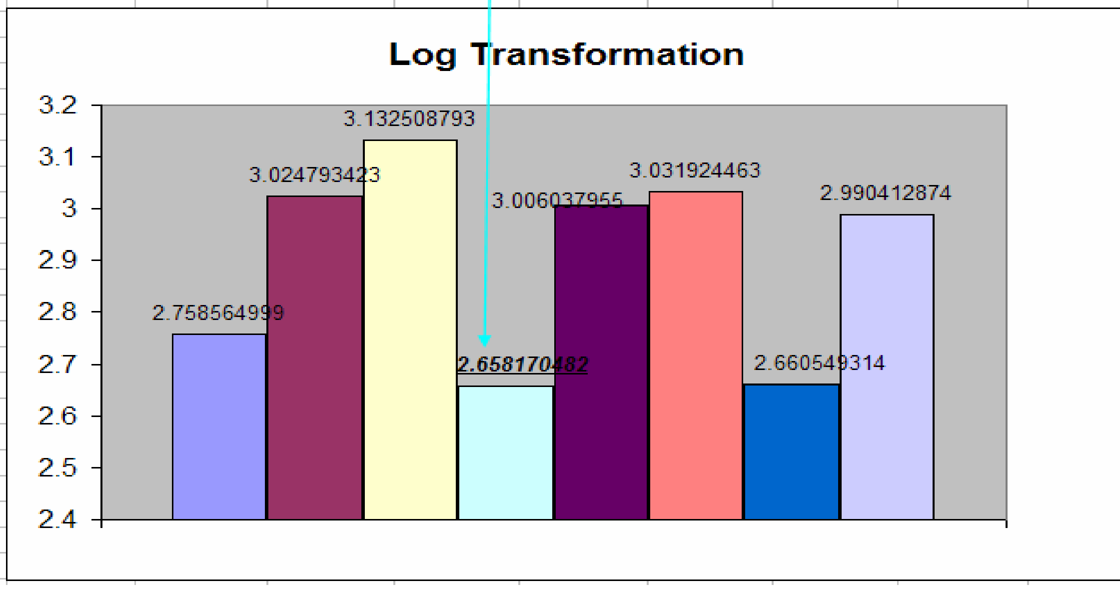

Figure 13: Log Transformation data depicting lens health as a function of opacity. The lower the value, the healthier the lens.

Conclusion

The BI-9000AT Program was enhanced by SRDAM software. The SRDAM Program analyzed the data analysis technique more efficiently than the BI-90000AT Program alone. This program scanned multiple data files and allowed the application of statistical functions to show the relationship between complications seen in diabetes, such as blindness due to cataracts, using the sand rat as a model.

Future Research

However, in the future development of more feasible modules is important. In future, a complete graphics packet could be developed to load data file (.data extension) and perform al necessary statistical function through GUI (Graphical User Interface) and eliminate the need of BI-9000AT program by improving SRDAM software.

References

[1] M. Kore. Diabetes Mellitus in the Era of Proteomics. *Molecular & Cellular Proteomics* 2: 399-404, 1993.

[2] A.J. Bron., J. Sparrow, N.A.P. Brown., J. J. Harding and R. Blakytny. The Lens in Diabetes. *Eye* 7: 260-275, 1993.
[3] A. Foster. Cataract- A Global Perspective: Outcome and Outlay. *Eye* 3:449-453.
[4] J. H. Adler., L. Yanko., R. Salitenik, & A. Browne. "Study of Hyperglycemia and Cataracts within a Standardized Breeding Program for Sand Rats (*Psammomys obesus*)," In: Lessons from Animal Diabetes, Libbe, London. 500- 507,1984
[5] G. Marquie., J. Duhalt., P. Hadjiisky., P. Petkov & H. Bouissou. Diabetes Mellitus in Sand Rats (*Psammomys obesus*): microangiopathy during development of the diabetic syndrome. *Cell Mol Phys* 37(6), 652-667, 1991.
[6] J. H. Adler., C. Roderig, & A. Gutman. Breeding Sand Rats (*Psammomys obesus*) with a Diabetic Predisposition for Laboratory Investigation. Refuah. 33(1). 1-8, 1976.
[7] D.B. Hackel., K.Schmidt-Nielsen., H.B. Haines, & E. Mikat. Diabetes Mellitus in Sand Rats (*Psammomys obesus*) *Laboratory Investigation* 14(2): 200-207, 1995.
[8] J. Shulman, "Cataracts: The Complete Guide from Diagnostics to Recovery for Patients and Families," St. Martins Press, New York, 1993.
[9] G. Frenkel, P.F. Kraicer. Metabolic Pattern of Sand Rats, *Psammomys obes*us. *Life Science* 11: 209-222. 1972
[10] V.M., Chenault, M..N. Ediger, R.R. Ansari.. In Vivo Assessment of Diabetic Lenses Using Dynamic Light Scattering: *Diabetes Technology and Therapeutics*. 4(5), 651-659, 2002.
[11] Instruction Manual for BI-9000AT Digital Autocorrelator ver 1.25, Brookhaven Instruments Corporation, Holtsville: New York. 1998.
[12] H. S. Dhadwal., R.R. Ansari., M.A DellaVecchia and S. Dubin, "Fiber optic system for in-vivo sizing of protein in animal lenses." SPIE Proceedings Vol. 2393: Ophthalmic Technologies V: 227-236, 1995.
[13] R.R. Ansari, & K. I. Suh. "Three-dimensional scanning of eye lens and vitreous using a newly developed dynamic light scattering probe" SPJE Proceedings Vol. 2673: Ophthalmic Technologies VI, 12-20, 1996.
[14] R.R. Ansari., K. I. Sub., M. A. Della Vecchia, & S. Dubin. "Ophthalmic diagnostics using a new dynamic light scattering fiber optic probe", SPIE Proceedings Vol. 2632, Lasers in Ophthalmology III, 62-72, 1996.
[15] R.R. Ansari., H. S. Dhadwal., M. C. Campbell, & M. A. Della Vecchia, "Fiber optic sensor for ophthalmic refractive diagnostics", SPIE Proceedings. Vol. 1648: Fiber Optic Medical and Fluorescent Sensors and Applications, 83-105, 1992.
[16] V.M. Chenault. "Use of Psammomys obesus(Fat Sand Rat) as a Cross Cutting Animal Model of Risk-Diabetes to Devices", Presentation @ Talladega College, AL. Program C: Biological Risk Assessment ID# RP 1023. Nov. 4, 2005.
[17] World-Wide Web URL http://www.spss.com/, Last Accessed October 10, 2005, Chicago, Illinois.

Biographies

SYED RAZA is Chair and Associate Professor of Computer Science, Department of Mathematics and Computer Science at Talladega College in Talladega, Alabama. He earned his

B.Sc. (Science, 1993), M.S. (Computer Science, 1998) degrees from University of the Punjab. He earned his M.S (Computer Science, 2004) and is completing a Ph.D., ABD (Computer Information Systems, 2006) from Nova Southeastern University in Florida. He teaches a variety of courses to undergraduate computer science major and minor students. His research is in RFID, Knowledge Management and Data Analysis techniques such as Data Mining and its application.

MARC A. WILSON is currently a student intern at the Food and Drug Administration (FDA), Center for Devices and Radiological Health in Rockville, MD. He earned a B.S. (Computer Science, 2006) from Talladega College and is entering graduate school in the fall of 2006.

LEONARD S. COLE is Director of Office of Sponsored Programs at Talladega College. He is a Professor of Biology in the Natural Sciences and directs three projects funded by NSF. He earned his B.A. (Biology, 1963) Talladega College, Talladega, AL; MST (Biology Education, 1968) Southern University, Baton Rouge, LA; Ph.D. (Cell and Developmental Biology, 1979) Atlanta University, Atlanta, GA.

SHIKHA GOLA earned a B.S. (Biomedical Engineering, 2006) from Case Western Reserve University in Ohio. She participated on this project during a student internship with the FDA in 2005 and 2006. She is a Regulatory Associate with Boston Scientific, Inc. in Minneapolis, MN.

V. MICHELLE CHENAULT is Associate Director of the Medical Device Fellowship Program with the FDA in Rockville, MD. She earned her B.A. (Liberal Arts- Major: Biology, Minor: Chemistry and Medical Technology certification, 1975) from Valparaiso University in Indiana, M.S. (Education, 1978) from Central Michigan University and Ph.D. (Molecular and Cellular Pathobiology, 1991) from Wake Forest University in North Carolina. She is a principal investigator and adjunct faculty at Uniformed Services University of the Health Sciences in Bethesda, MD and Talladega College and conducts research studies using *Psammomys obesus*.

Funded by: FDA Grant# G570MY; NSF Grant # HRD-0207965; NIH/NICHD Grant # HD046389.

Acknowledgements: Julia B. Anderson, Ph.D. Assistant to the Deputy Director, NIH/NICHD; Ms. Carol Schnupp, MS, USUHS Research Associate.