
Dominik Ślęzak · Tugh
Wai-Chi Fang Xiaofen
Tai-hoon Kim (Eds.)
Communications in Computer a
Bio-Scienc
and Bio-Te

Communications
in Computer and Information Science 57

Dominik Ślęzak Tughrul Arslan
Wai-Chi Fang Xiaofeng Song
Tai-hoon Kim (Eds.)

Bio-Science and Bio-Technology

International Conference, BSBT 2009
Held as Part of the Future Generation
Information Technology Conference, FGIT 2009
Jeju Island, Korea, December 10-12, 2009
Proceedings

Springer

Volume Editors

Dominik Ślęzak
University of Warsaw & Infobright Inc., Warsaw, Poland
E-mail: slezak@infobright.com

Tughrul Arslan
University of Edinburgh, Edinburgh, UK
E-mail: tughrul.arslan@ee.ed.ac.uk

Wai-Chi Fang
National Chiao Tung University, Hsinchu, Taiwan
E-mail: wfang@mail.nctu.edu.tw

Xiaofeng Song
Shanghai University, Shanghai, China
E-mail: mail2song@163.com

Tai-hoon Kim
Hannam University, Daejeon, South Korea
E-mail: taihoonn@hnu.kr

Library of Congress Control Number: 2009939455

CR Subject Classification (1998): J.3, I.2.1, H.2, H.2.8, K.8.1, J.1

ISSN 1865-0929
ISBN-10 3-642-10615-3 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-10615-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12807642 06/3180 5 4 3 2 1 0

Foreword

As future generation information technology (FGIT) becomes specialized and fragmented, it is easy to lose sight that many topics in FGIT have common threads and, because of this, advances in one discipline may be transmitted to others. Presentation of recent results obtained in different disciplines encourages this interchange for the advancement of FGIT as a whole. Of particular interest are hybrid solutions that combine ideas taken from multiple disciplines in order to achieve something more significant than the sum of the individual parts. Through such hybrid philosophy, a new principle can be discovered, which has the propensity to propagate throughout multifaceted disciplines.

FGIT 2009 was the first mega-conference that attempted to follow the above idea of hybridization in FGIT in a form of multiple events related to particular disciplines of IT, conducted by separate scientific committees, but coordinated in order to expose the most important contributions. It included the following international conferences: Advanced Software Engineering and Its Applications (ASEA), Bio-Science and Bio-Technology (BSBT), Control and Automation (CA), Database Theory and Application (DTA), Disaster Recovery and Business Continuity (DRBC; published independently), Future Generation Communication and Networking (FGCN) that was combined with Advanced Communication and Networking (ACN), Grid and Distributed Computing (GDC), Multimedia, Computer Graphics and Broadcasting (MulGraB), Security Technology (SecTech), Signal Processing, Image Processing and Pattern Recognition (SIP), and u- and e-Service, Science and Technology (UNESST).

We acknowledge the great effort of all the Chairs and the members of advisory boards and Program Committees of the above-listed events, who selected 28% of over 1,050 submissions, following a rigorous peer-review process. Special thanks go to the following organizations supporting FGIT 2009: ECSIS, Korean Institute of Information Technology, Australian Computer Society, SERSC, Springer LNCS/CCIS, COEIA, ICC Jeju, ISEP/IPP, GECAD, PoDIT, Business Community Partnership, Brno University of Technology, KISA, K-NBTC and National Taipei University of Education.

We are very grateful to the following speakers who accepted our invitation and helped to meet the objectives of FGIT 2009: Ruay-Shiung Chang (National Dong Hwa University, Taiwan), Jack Dongarra (University of Tennessee, USA), Xiaohua (Tony) Hu (Drexel University, USA), Irwin King (Chinese University of Hong Kong, Hong Kong), Carlos Ramos (Polytechnic of Porto, Portugal), Timothy K. Shih (Asia University, Taiwan), Peter M.A. Sloot (University of Amsterdam, The Netherlands), Kyu-Young Whang (KAIST, South Korea), and Stephen S. Yau (Arizona State University, USA).

We would also like to thank Rosslin John Robles, Maricel O. Balitanas, Farkhod Alisherov Alisherovish, and Feruza Sattarova Yusfovna – graduate students of Hannam University who helped in editing the FGIT 2009 material with a great passion.

October 2009

Young-hoon Lee
Tai-hoon Kim
Wai-chi Fang
Dominik Ślęzak

Preface

We would like to welcome to the proceedings of the 2009 International Conference on Bio-Science and Bio-Technology (BSBT 2009), organized as part of the 2009 International Mega-Conference on Future Generation Information Technology (FGIT 2009), held during December 10–12, 2009, at the International Convention Center Jeju, Jeju Island, South Korea.

BSBT 2009 focused on various aspects of advances in bio-science and bio-technology with computational sciences, mathematics and information technology. It provided a chance for academic and industry professionals to discuss recent progress in the related areas. We expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject.

We would like to acknowledge the great effort of all the Chairs and members of the Program Committee. Out of around 70 submissions to BSBT 2009, we accepted 22 papers to be included in the proceedings and presented during the conference. This gives roughly a 30% acceptance ratio. Two of the papers accepted for BSBT 2009 were published in the special FGIT 2009 volume, LNCS 5899, by Springer. The remaining 20 accepted papers can be found in this CCIS volume.

We would like to express our gratitude to all of the authors of submitted papers and to all of the attendees, for their contributions and participation. We believe in the need for continuing this undertaking in the future.

We would also like to acknowledge KuoYuan Hwa (National Taipei University of Technology, Taiwan) and Chien-Yeh Hsu (Taipei Medical University, Taiwan) for organizing the special session on Advanced Information and Computational Technology for Biomedical Research.

Once more, we thank all the organizations and individuals who supported FGIT 2009 as a whole and, in particular, helped in the success of BSBT 2009.

October 2009

Dominik Ślęzak
Tughrul Arslan
Wai-chi Fang
Xiaofeng Song
Tai-hoon Kim

Organization

Organizing Committee

General Chairs	Tughrul Arslan (Edinburgh University, UK) Wai-chi Fang (National Chiao Tung University, Taiwan)
Advisory Board	Saman Halgamuge (University of Melbourne, Australia) Joseph Kolibal (University of Southern Mississippi, USA) Philip Maini (University of Oxford, UK) Byoung-Tak Zhang (Seoul National University, Korea)
Program Chairs	Xiaofeng Song (NUAA, Nanjing, China) Tai-hoon Kim (Hannam University, Korea)

Program Committee

Asai Asaithambi
A.Q.K. Rajpoot
Adrian Stoica
Ajay Kumar
Arun Ross
Bob McKay
Carlos Juiz
Cesare Alippi
Dana Lodrova
Davide Anguita
Dong-Yup Lee
Emilio Corchado
Farzin Deravi
Francisco Herrera
George A. Gravvanis
Hujun Yin
Janusz Kacprzyk
Jason T.L. Wang
Javier Ortega-Garcia
Jim Torresen
Jongwook Woo
J.A. Ferreira Costa
José Manuel Molina
Juan Manuel Corchado
Kayvan Najarian
Kenji Mizuguchi
Kevin Daimi
Li Xiaoli
Liangjiang Wang
Lusheng Wang
Martin Drahansky
Matthias Dehmer
Meena K. Sakharkar
Michael E. Schuckers
Pong C. Yuen
QingZhong Liu
R. Ponalagusamy
Rattikorn Hewett
Saman Halgamuge
Stan Z. Li
Stephen Cameron
Suash Deb
Tatsuya Akutsu
Tommaso Mazza
Waleed Abdullah
Wei Zhong
Witold Pedrycz
Xiaohua (Tony) Hu
Yaoqi Zhou
Yong Shi
Yu Zheng
Zhenan Sun
Zizhong Chen

Table of Contents

Advanced Information and Computational Technology for Biomedical Researches (Special Session)

The Weighting Analysis of Influence Factor in Clinical Skin Physiology Assessment via Rough Set Method

Hui-Yi Liang[1], Ya-Ting Lee[2], Mei-Li You[3], and Kun-Li Wen[4]

[1] Department of Applied Foreign Language, Chienkuo Technology University, Changhua, Taiwan
[2] Department of Beauty Science, Chienkuo Technology University, Changhua, Taiwan
[3] Department of General Education, Chienkuo Technology University, Changhua, Taiwan
[4] Department of Electrical Engineering (GSRC), Chienkuo Technology University Changhua, Taiwan
klw@ctu.edu.tw

Abstract. The main purpose of this paper is to study the weighting analysis of clinical skin physiology assessment. First of all, we analyze skin's physiological factors, which include four factors: Cutometer, PH value, Sebumeter and Mx16. A detailed description of each influence factors is offered and based on the analysis of 61 experimental objects' testing numerical values. Then, using rough set method in soft computing theory and the weighting model, we derived the factors' weighting and relational of skin physical system from the skin characteristic analysis of diverse-aged research objects. This paper also apply Matlab to develop a complete human-machine interface type of toolbox in order to support the calculation and verification the huge data. Finally, some further suggestions are indicated for the research in the future.

Keywords: Clinical skin physiology assessment, Weighting, Rough set, Matlab.

1 Introduction

The condition of skin cannot be seen by naked eyes. It needs special equipment to test the physical properties of skin before determining its conditions. Clinically, there are various instruments for testing skin, though most simply analyze the general characteristics without diagnosing it in detail and further analyzing it according to age differences. As the largest human organ, besides protecting the human body, skin manifests human physical and mental conditions. This study attempts to analyze the weighting of factors in skin physiology assessment based on the physical conditions of various clinical cases, including the elasticity, pH value, sebum, color, moisture, and hydrating capacity of skin[1,2]. The significant analysis in rough set theory is

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 1–8, 2009.

applied to analyze the weighting of various factors in skin physiology assessment because this mathematical model fits the situation. The rough set is often used in applied to practical analysis, operated manually in analyses and calculations, programmed with general languages even as a computer utility. For example, many Chinese version utilities and other software utility researches using the rough set are programmed with C^{++} and Visual Basic. As the rough set theory has been developed for over 20 years[3~5], applied to analyses in Taiwan for over a decade, and has become a popular research topic[6], this paper will integrate rough set method in software to develop corresponding and practical computer utilities with the Matlab[7~9]. Next, data will be collected from subjects of different skin quality selected clinically in order to develop a practical system based on the results of skin factor weighting analysis as a reference for cosmetology and related researches. Though there were some skin-related studies using the rough set theory, all are qualitative researches[10,11].

Therefore, this is the first quantitative study on skin physiology assessment analyzing factor weight with the rough set theory in a software mathematical model. In this paper, first, in section 2, we list the whole mathematical foundation of rough set in detail. In section 3, the basic concept of clinical skin physiology assessment and its measurement method are presented. Section 4 consists of empirical analysis where actual data was substituted into the mathematical model to derive the needed data. The final section of this study consists of a conclusion and recommendations for future research.

2 Paper Preparation

In this section, we only simply introduce the basic concept of rough set[6].

2.1 Basic Relationship

1. Information system, IS
   $IS = (U, A)$ is called information system, where $U = \{x_1, x_2, x_3, ..., x_n\}$ is the universe finite set of object, and $A = \{a_1, a_2, ..., a_m\}$ is the set of attribute.
2. Information function
   If exist a mapping $f_a : U \to V_a$, then V_a is the set of value of a, call the domain of attribute a.
3. Indiscernibility relation: An indiscernibility relation is defined as For any x_i and x_j, if x_i is identical to x_j, then x_i and x_j have all the same properties.

2.2 Calculation Method

1. If $A \subseteq U$, then the lower approximation is defined as

$$\underline{R}(A) = \{x \in U \mid [x]_R \subseteq A\} = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \subseteq A\} \tag{11}$$

and upper approximation is defined as

$$\overline{R}(A) = \{x \in U \mid [x]_R \cap A \neq \phi\} = \bigcup \{[x]_R \in \frac{U}{R} \mid [x]_R \cap A \neq \phi\} \tag{12}$$

2. Positive, negative and boundary:
Base on the mentioned above, the positive and negative are defined as

$$pos_R(X) = \underline{R}(X),\ neg_R(X) = U - \overline{R}(X) \tag{13}$$

also the boundary of A is defined as

$$bn_R(A) = \underline{R}(A) - \overline{R}(A) \tag{14}$$

3. The dependents of attributes

$$\gamma_c(D) = \frac{|posc(D)|}{U} \tag{15}$$

Means under $a \in C$, the ratio in the whole set

4. The significant value of attributes:
In $S = (U, C \cup D, V, f)$ system, under $a \in C$, the significant value of attributes is defined as

$$\sigma_{(C,D)}(a) = \frac{\gamma_c(D) - \gamma_{c-\{a\}}(D)}{\gamma_c(D)} \tag{16}$$

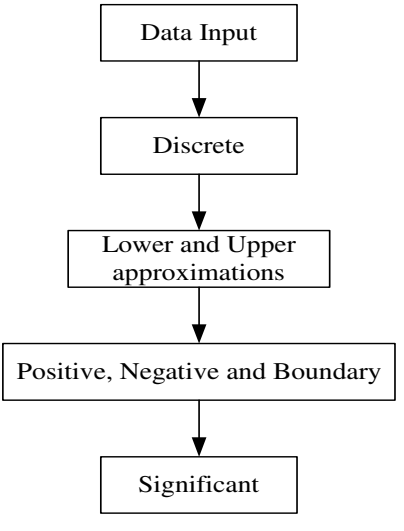


Fig. 1. The operation of weighting in rough set method

3 Clinical Skin Physiology Assessment

3.1 Clinical Skin Physiology Assessment

The application and analysis of clinical skin physiology assessment are as follows

1. Testing conditions: 20±2°C and RH 45±5%. As the best results can be obtained in such conditions, it is necessary to create such an environment for running the test.
2. Subjects: 50-70 subjects will be selected, and the coverage of subjects and testing results will be the data for the statistical analysis.
3. Age and gender of subjects: over 18-52 years of age, females.
4. Part of skin test: As the keratin content, moisture content, sebum secretion, and skin pigment distribution vary at different parts of the skin, this study selects the cheeks as the location of testing conducted three times.
5. Cleaning of the part of test: Subjects will wash their cheeks with non-alcoholic lotion or cleansing lotion 3 hours before the test is conducted to let the skin recover its original condition.
6. Skin type: The subjects with dry and oily skin were selected on an equal basis after testing for three times. Those with skin irritation or allergy will be eliminated from the test.

3.2 Skin Physiology Analyzers

1. Cutometer: Using the ultrasonic theory, as the elastic fiber tissues crisscross, it needs to measure the same point of measurement from four different angles and calculate the mean elasticity of these four measurements.
2. PC Skin-pH-Meter 905: Improved from industrial pH testing rod, it judges the results with particular software or an LCD panel.
3. Sebumeter-810: It uses a special type of translucent papers that become transparent after contacting sebum, and the greater the amount of sebum, the more transparent the paper will become. Clean test papers will first be scanned once in the infrared sensing area before attaching them to the testing area for a particular time specified by the instrument. Then, the test papers absorbed the sebum will be scanned again to measure the amount of sebum.
4. Mexameter MX16: Using the spectral reflecting theory: It measures the LAB value with full spectrum or the melanin and hemoglobin counts at a particular wavelength.
5. Corneometer: Using the human conductivity theory: An electric current will be administered on the keratin. By calculating the impedance, the hydrating capacity of skin can be obtained.
6. Tewameter: This uses an ultrasensitive microchip to capture the water molecules diffused from the skin to measure the water loss speed of skin.

3.3 Evaluation Factors

This study assesses human skin physiology in terms of five factors: skin elasticity, skin pH value, sebum secretion, skin pigment, and age as described in the mentioned above.

4 Results Analysis

A total of 61 female subjects were selected from central Taiwan aged between 18 and 52 years of age. The weighting model was applied to convert skin physiological variables into factors for analysis in order to calculate the weighting of various factors.

The original data are shown in Table 1, and the discrete data are shown in Table 2, where cutometer takes large the better; PH value takes nominal better; sebumeter takes small the better; Mx16 takes large the better, and age takes large the better, the discrete are take five equal interval.

Based on mathematical analysis, the significant of various factors on the skin of dispersed subjects at different ages was calculated with equations (11) to equation (16). The value is the level of attribute significant that will affect the weighting of factor outcomes as shown in Table 3.

Table 1. The original measurement data

No.	Cutometer	PH	Sebumeter	Mx16	Age
1	0.94	4.55	120	166	19
2	0.53	5.42	50	320	52
3	0.62	5.13	90	326	48
4	0.71	5.22	130	250	40
5	0.95	5.33	257	168	20
6	0.92	4.89	220	188	21
7	0.91	5.31	256	367	19
8	0.92	5.43	187	245	20
9	0.9	4.80	290	260	21
10	0.93	5.44	210	157	24
11	0.92	4.86	188	268	28
12	0.88	5.33	185	156	33
13	0.87	4.88	82	199	39
14	0.89	5.28	150	268	20
15	0.87	4.86	121	150	24
⋮	⋮	⋮	⋮	⋮	⋮
49	0.54	4.65	188	255	25
50	0.67	6.66	89	228	23
51	0.86	5.70	143	167	23
52	0.88	5.38	157	198	24
53	0.76	5.88	140	217	25
54	0.78	5.95	114	142	24
55	0.82	6.15	227	192	22
56	0.66	5.78	120	152	25
57	0.72	5.72	89	160	23
58	0.62	5.90	129	160	24
59	0.83	5.68	256	227	28
60	0.62	5.79	124	232	23
61	0.75	5.45	156	142	24

Table 2. The measurement data after discrete

No.	Cutometer	PH	Sebumeter	Mx16	Age
1	5	1	4	1	1
2	1	3	5	4	3
3	2	2	5	4	3
4	3	2	4	2	2
5	5	3	1	1	2
6	5	2	2	1	2
7	5	3	1	5	1
8	5	3	3	2	1
9	5	1	1	3	2
10	5	3	2	1	1
11	5	2	3	3	1
12	4	3	3	1	1
13	4	2	5	2	3
14	5	3	3	3	1
15	4	2	4	1	1
⋮	⋮	⋮	⋮	⋮	⋮
49	1	1	3	3	1
50	2	1	5	2	2
51	4	3	4	1	1
52	4	3	3	2	2
53	3	2	4	2	1
54	3	2	4	1	5
55	4	2	2	1	2
56	2	2	4	1	1
57	3	3	5	1	2
58	2	2	4	1	1
59	4	3	1	2	2
60	2	2	4	2	1
61	3	3	3	1	1

Table 3. The analysis results

Factor	Cutometer	PH	Sebumeter	Mx16
Weighting	0.2830	0.1698	0.4151	0.3962
Rank	III	IV	I	II

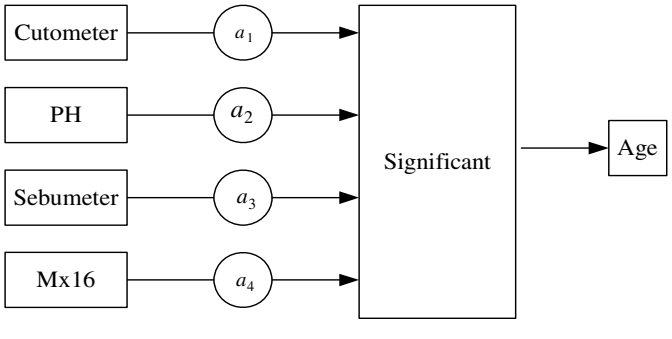


Fig. 2. The analysis structure

Because of the amount of data are enormous, therefore, the computer toolbox is also used to analyze and verify [8], as shown in Fig. 2.

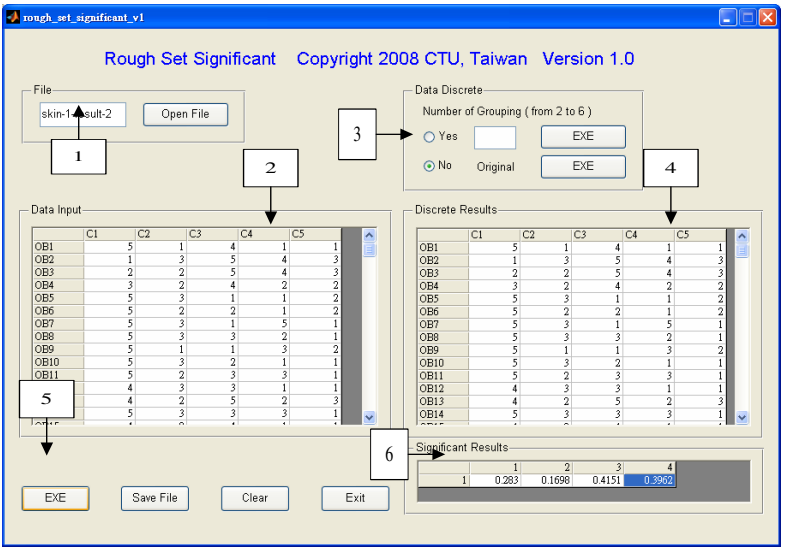


Fig. 2. The screen of toolbox by using significant in rough set method

where:

1. Input the inspected data (Excel Type)
2. The original data is displayed
3. Choose discrete method (use original data)
4. The data after discrete
5. Press the EXE icon
6. The Significant is appeared

5 Conclusions and Discussions

Conventional statistical analyses were applied to most studies on skin physiology assessment. As they need a large amount of data, it is clinically difficult to fulfill relevant requirements. As a result, results of these studies were unable to reflect the reality. The significant rough set method that used in this study can overcome these problems because the model developed in this study is a non-function serial model that is easy to use and does not need to conform to the typical distribution regularities. The practical measuring system developed with the rough set model in this study analyzes the weighting, and relational of skin physiological factors based on the skin characteristics of subjects at different ages, and delivers results that match the conventional concepts. Therefore, it can be a reliable reference for related researches. Besides, Matlab has been applied to develop a computer utility for this analysis model in order to develop an integrated computer analysis utility for processing a large amount of data.

To sum up, the rough set theory that used in this study is one of the software calculation methods. Therefore, in addition to including other related influence factors, further studies can increase the amount of data in order to enhance validity. Furthermore, other software calculations, such as the fuzzy, ANN, and grey system theory, can be integrated to the rough set theory to enhance the reliability of results.

References

1. Black, D., Pozo, A., Del, G.Y.: Seasonal and anatomical variations in the surface state of the stratum corneum. J. Investigative Dermatology 108, 824 (1997)
2. Sator, P.G., Schmidt, J.B., Hönigsmann, H.: Comparison of epidermal hydration and skin surface lipids in healthy individuals and in patients with atopic dermatitis. J. the American Academy of Dermatology 48, 352–358 (2003)
3. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11(5), 341–356 (1982)
4. Pawlak, Z.: Rough sets, theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991)
5. Pawlak, Z.: Rough sets approach to multi-attribute decision analysis. European Journal of Operational research 72, 443–459 (1994)
6. Wen, K.L., Nagai, M.T., Chang, T.C., Wen, H.C.: An introduction to rough set and application. Wunam Published, Taipei (2008)
7. Lee, S.H.: Matlab 7.x: Interface development & complier method. Kingsinformation Publisher, Taipei (2006)
8. Wen, K.L., Changchien, S.K.: The weighting analysis of influence factors in gas breakdown via rough set and GM(h,N). J. Computers 3(11), 17–24 (2008)
9. Wen, K.L.: Apply MATLAB in grey system theory. CHWA Publisher, Taipei (2006)
10. Wissing, S.A., Müller, R.H.: The influence of solid lipid nanoparticles on skin hydration and viscoelasticity- in vivo study. European Journal of Pharmaceutics and Biopharmaceutics 56, 67–72 (2003)
11. Tupker, R.A., Schuur, J., Coenraads, P.J.: Irritancy of antiseptics tested by repeated open exposures on the human skin, evaluated by non-invasive methods. Occupational Health and Industrial Medicine 38, 83–84 (1998)

Identifying Candidate Disease Gene GAD2 for Obesity by Computational Gene Prioritization Tool ENDEAVOUR

Huanping Zhang, Xiaofeng Song*, and Huinan Wang

Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China
xfsong@nuaa.edu.cn

Abstract. Identifying potential disease genes for complex disease is very important for understanding disease pathogenesis and providing preventive measures. The computational gene prioritization tools provide a promising and effective way for disease gene identification. Gene GAD2 on Chromosome 10p12 is a disputing candidate disease gene for obesity because of the contradictory conclusions obtained by association studies of different researchers. In this paper, a computational tool ENDEAVOUR is used to verify the probability of gene GAD2 being an obesity candidate disease gene. ENDEAVOUR evaluates each candidate gene based on its similarity with the training genes (known disease genes). The high prioritization of gene GAD2 in the computational results means that the gene GAD2 has high probability to be a disease gene for obesity.

Keywords: Disease Gene, Obesity, ENDEAVOUR.

1 Introduction

Finding potential human disease genes from the whole genome for complex disease is very important to understand disease pathogenesis and improve clinic treatment. But identifying genes underlying complex disease is not easy because complex disease generally involves multiple contributing genes [1-2]. Traditional experimental methods such as association study and linkage analysis are labor-intensive; meanwhile analyzing the increasingly large candidate gene sets generated by experimental methods is time consuming. So some computational disease gene prioritization tools are proposed, which employ huge amounts of biology data from databases such as sequence data, biological information, functional information and expression data etc. The increasing wealth of such genomic data available in public databases enables these tools more comprehensive and powerful. All these disease gene prioritization tools such as POCUS[3], G2D[4], GeneWanderer[5], PROSPECTR[6], SUSPECTS[7], TOM[8], GFINDer[9], GeneSeeker[10], Endeavour[11,12] provide an effective and promising way to disease genes prediction. For example, Tiffin et al. reviewed seven

* Corresponding author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 9–16, 2009.

independent computational disease gene prioritization methods, and then apply them to the analysis of 9556 positional candidate genes for type 2 diabetes and obesity [13].

Obesity is a known risk factor for many chronic diseases and health conditions including type2 diabetes, hypertension, coronary heart disease, and certain forms of cancer etc [14]. In addition to the effects of environmental, psychological and other factors, susceptibility to obesity is strongly influenced by multiple genetic factors [15]. It is not easy to identify causative genes for this complex disease. Many researchers have done lots of work on finding candidate disease genes for this complex obesity disease [16-18].

Gene GAD2 on Chromosome 10p12 is proposed to be a disputing candidate gene for obesity. Boutin et al. have performed case-control study (575 morbidly obese and 646 control subjects) in four independent ethnic populations and concluded that the gene GAD2 encoding the glutamic acid decarboxylase enzyme (GAD65) is a promising candidate gene for human obesity [19, 20]. The results of Meyre et al. also confirmed the association between gene GAD2 promoter SNP and the risk for obesity, meanwhile the results indicated that gene GAD2 is a genetic link between birth weight and subsequent development of obesity in children [21].

However Swarbrick et al. attempted to replicate the important findings of Boutin et al. they performed family-based tests of association in 2,359 German Caucasian individuals from 693 nuclear families, the conflicting results indicated lack of support for the association between GAD2 polymorphisms and severe human obesity[22]. Tiwari et al. reviewed the genetic and functional arguments for and against the GAD2 gene as an influential gene for obesity [23]. Some researcher thought that their work does not provide enough evidence between the gene GAD2 with obesity [24].

Why different researches may show different results, the possible reasons maybe are ethnic differences, population stratification, behavioral and dietary differences, lack of statistical power to detect small to moderate effects, and poor biological plausibility; the number of subjects involved is also a factor for the statistical power of their tests [22,25].

So in order to further the mechanism of obesity and function of gene GAD2, the disease gene prioritization tool ENDEAVOUR based on bioinformatics methods is utilized to identify the probability of gene GAD2 being a candidate gene for obesity.

2 Methods

Among many disease gene prediction tools, the powerful and comprehensive computational gene prioritization tool ENDEAVOUR, developed by scientists in university of Leuven is used to verify the disputing gene GAD2 [11, 12].

2.1 ENDEAVOUR

ENDEAVOUR is based on assumption that genes involved in the same disease or the same biological pathway are more similar one to each other than to the rest of the genome. 'Similar' can mean sequences similarities, or sharing functional regulation motifs or that their proteins are physically interacting within cells. ENDEAVOUR considers that the candidate gene of the highest similarity with the training gene is the best candidate gene. Many researchers have made use of ENDEAVOUR to prioritize

Table 1. Data types utilized by ENDEAVOUR

Name	Data types	
ENDEAVOUR	Annotation	EnsemblEst; GeneOntology; Interpro; Swissprot; Kegg
	Interaction	Bind; BioGrid; Hprd; InNetDb; Intact; Mint; String
	Expression	SonEtAl; SuEtAl
	Precalculated	Ouzounis; Prospectr
	Sequence based data	Blast
	Literature data	Text
	Regulatory information	Motif; CisRegModule

candidate genes for diseases. ENDEAVOUR research group identified YPEL1 as a candidate DiGeorge syndrome (DGS) gene by their prioritization tool. Elbers et al. have used ENDEAVOUR and several other prioritization tools for disease gene identification by analyzing five susceptibility loci for type 2 diabetes and obesity, and 27 functional candidate genes have been pinpointed that are involved in eating behaviors, metabolism and inflammation[26]. Osoegawa et al. applied ENDEAVOUR to propose novel genes associated with cleft lip and cleft palate phenotypes [27]. Tzouvelekis et al. have used ENDEAVOUR to prioritize a list of genes differentially expressed in idiopathic pulmonary fibrosis [28].

The updated ENDEAVOUR provides a web-based user interface to make it more universally accessible (http://www.esat.kuleuven.be/endeavour/). ENDEAVOUR support gene prioritization for multiple species including three major model organisms: Mus musculus, Rattus norvegicus and Caenorhabditis elegans in addition to Homo sapiens. ENDEAVOUR makes use of multiple data sources: ontologies, annotations, protein–protein interactions, cis-regulatory information, gene expression data sets, sequence information and text-mining, Table1 lists the data types utilized by ENDEAVOUR.

The prioritization processes consist of four steps: the first step is to choose the species; the second is collecting the training genes (genes known to be connected with a disease or a biological process under study), information of training genes are retrieved from numerous data sources in order to build models. The third step is data sources selection, and each data source build a sub-model. In the fourth step, the models built in previous steps are used to score the candidates genes. For each data source, the candidate genes are ranked according to their similarities to the appropriate sub-model. A set of rankings (one per data-source) for each gene is obtained; then the set of rankings per data source are fused into a global ranking using order statistics algorithm. The overall rank indicates the similarity between the candidate genes with the training genes.

2.2 Training Genes

The training genes should be collected by users before prioritization processes. Training genes are genes known to be related with obesity. Totally 16 genes are collected. The gene symbols and other information of these training genes are list in Table 2.

Table 2. Training genes for obesity used by ENDEAVOUR

Gene symbol	Ensembl ID	Locus	Entrez Gene	Swissprot ID
LEP	ENSG00000174697	7q31	3952	P41159
PCSK1	ENSG00000175426	5q15-q21	5122	P29120
MC4R	ENSG00000166603	18q22	4160	P32245
NTRK2	ENSG00000148053	9q22.1	4915	Q16620
SDC3	ENSG00000162512	1pter-p22.3	9672	O75056
UCP3	ENSG00000175564	11q13.4	7352	P55916
MC3R	ENSG00000124089	20q13.2-q13.3	4159	P41968
UCP2	ENSG00000175567	11q13	7351	P55851
PPARG	ENSG00000132170	3p25	5468	Q15180
ADIPOQ/ACDC	ENSG00000181092	3q27-29	9370	Q15848
ADRB2	ENSG00000169252	5q31-5q32	154	Q9UCZ3
ADRB3	ENSG00000188778	8p11-8p12	155	P13945
LEPROT/LEPR	ENSG00000116678	1p31	54741/3953	O15243
NR3C1	ENSG00000113580	5q31-5q32	2908	P04150
APOA4	ENSG00000110244	11q23	337	P06727
LDLR	ENSG00000130164	19p13.3	3949	P01130

The first 9 genes including LEP, PCSK1, MC4R, NTRK2, SDC3, UCP3, MC3R, UCP2, PPARG are from OMIM database[29]. Five genes ACDC, ADRB2, ADRB3, LEPR, and NR3C1 are training genes in PROSPECTR for obesity [6, 13]. The last two genes APOA4 and LDLR are from the genetic association database, they are located in regions studied with more than one independent non-negative association study (http://geneticassociationdb.nih.gov/) [13].

2.3 Candidate Genes and Data Sources

Totally 35 candidate genes are selected including gene GAD2. Another candidate gene FTO for obesity is also included, which is reported that the biological function of FTO is related to body weight controlling [30]. The rest 33 candidate genes are from results by Tiffin et al. [13]. Tiffin et al. collected all possible likely candidate genes previously identified by linkage and association studies, and described in the biomedical literature. Nine sub-models are selected: GeneOntology, Swissprot, Blast, InNetDb, Intact, Mint, String, Prospectr, and Text. The candidate genes should have enough information for each selected data type. These models are then used to score the candidate genes and rank them according to their scores, and then the rankings per data source are fused into a global ranking by using order statistics.

3 Results by ENDEAVOUR

The results of ENDEAVOUR can be displayed in three forms. The first panel is Sprintplot, which presents a graphical overview of the prioritization results. The first column corresponds to the global ranking of the genes, while the others give the ranking for each selected data source. The top 16 genes are attributed a random background color. Candidate genes in red color achieved a maximum dissimilarity

Global prioritization	GO	Swissprot	Blast	InNetDb	Intact	Mint	String	Prospectr	Text
GAD2	ZP2	FTO	ZNF579	GAD2	GAD2	GAD2	GAD2	ZZEF1	GAD2
ZP2	GAD2	ZP2	ZNF83	ZNRF4	ZNRF4	ZNRF4	ZNF652	GAD2	IKZF3
ZZEF1	ZNF8	GAD2	ZZZ3	ZP2	ZP2	ZP2	IKZF3	IKZF3	ZNF83
IKZF3	ZNHIT1	ZNRF4	ZBTB47	ZNF579	ZNF579	ZNF579	ZP2	ZNRF4	ZNF580
ZNF655	ZNRF4	ZZEF1	ZNF607	ZNF83	ZNF83	ZNF83	ZNF76	ZNRF2	ZNF84
FTO	ZZEF1	ZNF655	ZNF582	ZZZ3	ZZZ3	ZZZ3	ZNF655	ZBTB47	FTO
ZNRF4	IKZF3	ZNF577	FTO	ZBTB47	ZBTB47	ZBTB47	ZNF8	ZNHIT1	ZNHIT1
ZNHIT1	ZNF655	ZNF582	ZP2	ZNF607	ZNF607	ZNF607	ZZEF1	ZNF652	ZNF646
ZBTB47	ZNF594	ZNF324B	GAD2	ZNF582	ZNF582	ZNF582	ZNF646	ZNF76	ZNF655
ZNF83	ZNF652	ZNF79	ZNRF4	FTO	FTO	FTO	ZNF582	FTO	ZZEF1
ZNF76	ZNF619	ZNF76	ZZEF1	ZZEF1	ZZEF1	ZZEF1	ZNRD1	ZNRD1	ZNF75
ZNF8	ZZZ3	ZBTB47	ZNF655	ZNF655	ZNF655	ZNF655	ZNF580	ZP2	ZNRD1
ZNF579	ZNRF2	ZNRF2	ZNF577	ZNF577	ZNF577	ZNF577	ZNRF4	ZZZ3	ZNF606
ZNF646	ZNF580	ZNHIT1	ZNF324B	ZNF324B	ZNF324B	ZNF324B	ZNF579	ZNF579	ZNF76
ZNF582	ZNRD1	ZZZ3	ZNF79	ZNF79	ZNF79	ZNF79	ZNF83	ZNF655	ZNF8
ZNF580	ZNF577	ZNF8	ZNF76	ZNF76	ZNF76	ZNF76	ZZZ3	ZNF646	ZNF587
ZNF585A	ZSCAN1	ZNF579	ZNRF2	ZNRF2	ZNRF2	ZNRF2	ZBTB47	ZNF585A	ZNF79
ZNF84	ZNF582	IKZF3	ZNHIT1	ZNHIT1	ZNHIT1	ZNHIT1	ZNF607	ZNF607	ZNF585A
ZZZ3	ZNF324B	ZNF652	ZNF8	ZNF8	ZNF8	ZNF8	FTO	ZNF8	ZNF607
ZNF79	ZNF579	ZNF619	IKZF3	IKZF3	IKZF3	IKZF3	ZNF577	ZNF582	ZNF582
ZNF324B	ZNF624	ZNF580	ZNF652	ZNF652	ZNF652	ZNF652	ZNF324B	ZNF580	ZNF324B
ZNF75	ZNF607	ZSCAN1	ZNF619	ZNF619	ZNF619	ZNF619	ZNF79	ZNF83	ZSCAN1
ZNF587	ZNF587	ZNF624	ZNF580	ZNF580	ZNF580	ZNF580	ZNRF2	ZNF577	ZNF627
ZSCAN1	ZNF621	ZNF587	ZSCAN1	ZSCAN1	ZSCAN1	ZSCAN1	ZNHIT1	ZNF324B	ZNF624
ZNF652	ZNF75	ZNF621	ZNF624	ZNF624	ZNF624	ZNF624	ZNF619	ZNF79	ZNF594
ZNF624	ZNF83	ZNF75	ZNF587	ZNF587	ZNF587	ZNF587	ZSCAN1	ZSCAN1	ZP2
ZNRF2	ZNF79	ZNF83	ZNF621	ZNF621	ZNF621	ZNF621	ZNF624	ZNF624	ZNF579
ZNRD1	ZNF627	ZNF627	ZNF75	ZNF75	ZNF75	ZNF75	ZNF587	ZNF587	ZZZ3
ZNF607	ZNF76	ZNF646	ZNF627	ZNF627	ZNF627	ZNF627	ZNF621	ZNF621	ZNF577
ZNF577	ZNF646	ZNF84	ZNF646	ZNF646	ZNF646	ZNF646	ZNF75	ZNF75	ZNF621
ZNF621	ZNF606	ZNF585A	ZNF84	ZNF84	ZNF84	ZNF84	ZNF627	ZNF84	ZNRF2
ZNF627	ZNF84	ZNF594	ZNF585A	ZNF585A	ZNF585A	ZNF585A	ZNF84	ZNF606	ZNF652
ZNF619	ZNF585A	ZNRD1	ZNF594	ZNF594	ZNF594	ZNF594	ZNF585A	ZNF619	ZNRF4
ZNF606	~~FTO~~	ZNF607	ZNRD1	ZNRD1	ZNRD1	ZNRD1	ZNF594	ZNF627	~~ZBTB47~~
ZNF594	~~ZBTB47~~	ZNF606	ZNF606	ZNF606	ZNF606	ZNF606	ZNF606	ZNF594	~~ZNF619~~

Fig. 1. Sprintplot of 35 candidate genes ranked by ENDEAVOUR

score. Genes with name displayed in a line through font were not scored for the data source. The second panel is Table. Genes with a small score values share more similarity with training genes. The candidate genes are sorted according to the overall

ranking by default, or sorted according to any data-source. The third panel is Export, which allows saving the results, either in XML format or as CSV. The Sprintplot panel and the Table panel give the same global ranking order of candidate genes.

The results of these 35 candidate genes for obesity can be viewed by Sprintplot displayed in Fig.1. The first column corresponds to the global ranking of the 35 genes, while the following columns represent the ranking for each selected data source. Gene GAD2 in green color is of the highest global ranking, which means that gene GAD2 has more similarity with the training genes and of high probability to be a candidate disease gene for obesity. While another gene FTO has a lower prioritization. In the following sub-models: InNetDb, Intact, Mint, String, Text, gene GAD2 is also in the first rank.

4 Discussions

The computational results of ENDEAVOUR indicate that the gene GAD2 has high prioritization and high probability to be a candidate disease gene for obesity. These results are helpful to explain the contradictory conclusions obtained by other researchers. But it still needs support by sufficient genetic or biological evidence to elucidate genetic variation in gene GAD2 in the predisposition to severe obesity in humans.

The recent research results also confirmed that gene GAD2 is relevant to the obesity. The research group of Laval University have studied the association between GAD2 single-nucleotide polymorphisms (SNPs) and eating behaviors, dietary intake and obesity in subjects (n=873) from the Quebec Family Study (QFS).The results suggest a role for the GAD2 gene in determining food intake, eating behaviors and weight gain over time in women [31]. The research results of university of Pittsburgh indicated that single-nucleotide polymorphism (SNP) in the GAD2 gene (-243A-->G) was associated with increased BMI in late childhood and adolescence in the population of girls from western Pennsylvania [32]. The researchers from university of Oxford have evaluated gene GAD2 as a positional candidate for obesity [33].

At present people know little about obesity related genes, but future researches on biological mechanism of GAD2 and other potential obesity disease genes will resolve whether and how these genes influence body weight, and it is possible to identify those who may become obese later in life, then take preventative measures to ease the health burden associated with obesity.

References

1. Glazier, A.M., Nadeau, J.H., Aitman, T.J.: Finding genes that underlie complex traits. Science 298, 2345–2349 (2002)
2. Tabor, H.K., Risch, N.J., Myers, R.M.: candidate-gene approaches for studying complex genetic traits. Nat. Rev. Genet. 3, 391–397 (2002)
3. Turner, F.S., Clutterbuck, D.R., Semple, C.A.: POCUS: mining genomic sequence annotation to predict disease genes. Genome Biol. 4, R75 (2003)
4. Perez-Iratxeta, C., Wjst, M., Bork, P., Andrade, M.A.: G2D: a tool for mining genes associated with disease. BMC Genet. 6, 45 (2005)
5. Köhler, S., Bauer, S., Horn, D., Robinson, P.N.: Walking the interactome for prioritization of candidate disease genes. Am. J. Hum. Genet. 82, 949–958 (2008)

6. Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J., Pickard, B.S.: Speeding disease gene discovery by sequence based candidate prioritization. BMC Bioinformatics 6, 55 (2005)
7. Adie, E.A., Adams, R.R., Evans, K.L., Porteous, D.J., Pickard, B.S.: SUSPECTS: enabling fast and effective prioritization of positional candidates. Bioinformatics 22, 773–774 (2006)
8. Rossi, S., Masotti, D., Nardini, C., Bonora, E., Romeo, G., Macii, E., Benini, L., Volinia, S.: TOM: a web-based integrated approach for identification of candidate disease genes. Nucleic Acids Res. 34, W285–W292 (2006)
9. Masseroli, M., Galati, O., Pinciroli, F.: GFINDer: genetic disease and phenotype location statistical analysis and mining of dynamically annotated gene lists. Nucleic Acids Res. 33 (Web Server issue), W717–W723 (2005)
10. van Driel, M.A., Cuelenaere, K., Kemmeren, P.P., Leunissen, J.A., Brunner, H.G., Vriend, G.: GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. Nucleic Acids Res. 33, W758–W761 (2005)
11. Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B., Carmeliet, P., Moreau, Y.: Gene prioritization through genomic data fusion. Nature Biotechnology 24, 537–544 (2006)
12. Tranchevent, L.C., Barriot, R., Yu, S., Van Vooren, S., Van Loo, P., Coessens, B., De Moor, B., Aerts, S., Moreau, Y.: ENDEAVOUR update: a web resource for gene prioritization in multiple species. Nucleic Acids Res. 36(Web Server issue), W377–W384 (2008)
13. Tiffin, N., Adie, E., Turner, F., Brunner, H.G., van Driel, M.A., Oti, M., Lopez-Bigas, N., Ouzounis, C., Perez-Iratxeta, C., Andrade-Navarro, M.A., Adeyemo, A., Patti, M.E., Semple, C.A., Hide, W.: Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. Nucleic Acids Res. 34, 3067–3081 (2006)
14. Kopelman, P.G.: Obesity as a medical problem. Nature 404, 635–643 (2000)
15. Swarbrick, M.M., Vaisse, C.: Emerging trends in the search for genetic variants predisposing to human obesity. Curr. Opin. Clin. Nutr. Metab. Care 6, 369–375 (2003)
16. Bell, C.G., Walley, A.J., Froguel, P.: The genetics of human obesity. Nat. Rev. Genet. 6, 221–234 (2005)
17. Hinney, A., Ziegler, A., Oeffner, F., Wedewardt, C., Vogel, M., Wulftange, H., Geller, F., Stübing, K., Siegfried, W., Goldschmidt, H.P., Remschmidt, H., Hebebrand, J.: Independent confirmation of a major locus for obesity on chromosome 10. J. Clin. Endocrinol. Metab. 85, 2962–2965 (2000)
18. Saar, K., Geller, F., Rüschendorf, F., Reis, A., Friedel, S., Schäuble, N., Nürnberg, P., Siegfried, W., Goldschmidt, H.P., Schäfer, H., Ziegler, A., Remschmidt, H., Hinney, A., Hebebrand, J.: Genome scan for childhood and adolescent obesity in German families. Pediatrics 111, 321–327 (2003)
19. Boutin, P., Dina, C., Vasseur, F., Dubois, S., Corset, L., Séron, K., Bekris, L., Cabellon, J., Neve, B., Vasseur-Delannoy, V., Chikri, M., Charles, M.A., Clement, K., Lernmark, A., Froguel, P.: GAD2 on Chromosome 10p12 Is a Candidate Gene for Human Obesity. PLoS Biol. 1, e68 (2003)
20. Boutin, P., Froguel, P.: GAD2: A polygenic contribution to genetic susceptibility for common obesity? Pathol. Biol (Paris) 53, 305–307 (2005)
21. Meyre, D., Boutin, P., Tounian, A., Deweirder, M., Aout, M., Jouret, B., Heude, B., Weill, J., Tauber, M., Tounian, P., Froguel, P.: Is glutamate decarboxylase 2 (GAD2) a genetic link between low birth weight and subsequent development of obesity in children? J. Clin. Endocrinol. Metab. 90, 2384–2390 (2005)

22. Swarbrick, M.M., Waldenmaier, B., Pennacchio, L.A., Lind, D.L., Cavazos, M.M., Geller, F., Merriman, R., Ustaszewska, A., Malloy, M., Scherag, A., Hsueh, W.C., Rief, W., Mauvais-Jarvis, F., Pullinger, C.R., Kane, J.P., Dent, R., McPherson, R., Kwok, P.Y., Hinney, A., Hebebrand, J., Vaisse, C.: Lack of support for the association between GAD2 polymorphisms and severe human obesity. PLoS Biol. 3, e315 (2005)
23. Tiwari, H.K., Bouchard, L., Pérusse, L., Allison, D.B.: Is GAD2 on chromosome 10p12 a potential candidate gene for morbid obesity? Nutr. Rev. 63, 315–319 (2005)
24. Hunt, S.C., Xin, Y., Wu, L.L., Hopkins, P.N., Adams, T.D.: Lack of association of glutamate decarboxylase 2 gene polymorphisms with severe obesity in utah. Obesity 14, 650–655 (2006)
25. Cardon, L.R., Bell, J.I.: Association study designs for complex diseases. Nat. Rev. Genet. 2, 91–99 (2001)
26. Elbers, C.C., Onland-Moret, N.C., Franke, L., Niehoff, A.G., van der Schouw, Y.T., Wijmenga, C.: A strategy to search for common obesity and type 2 diabetes genes. Trends Endocrinol. Metab. 18, 19–26 (2007)
27. Osoegawa, K., Vessere, G.M., Utami, K.H., Mansilla, M.A., Johnson, M.K., Riley, B.M., L'Heureux, J., Pfundt, R., Staaf, J., van der Vliet, W.A., Lidral, A.C., Schoenmakers, E.F., Borg, A., Schutte, B.C., Lammer, E.J., Murray, J.C., de Jong, P.J.: Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. J. Med. Genet. 45, 81–86 (2008)
28. Tzouvelekis, A., Harokopos, V., Paparountas, T., Oikonomou, N., Chatziioannou, A., Vilaras, G., Tsiambas, E., Karameris, A., Bouros, D., Aidinis, V.: Comparative expression profiling in pulmonary fibrosis suggests a role of hypoxia-inducible factor-1alpha in disease pathogenesis. Am. J. Respir. Crit. Care Med. 176, 1108–1119 (2007)
29. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A.: Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 30, 52–55 (2002)
30. Loos, R.J., Bouchard, C.: FTO: the first gene contributing to common forms of human obesity. Obes. Rev. 9, 246–250 (2008)
31. Choquette, A.C., Lemieux, S., Tremblay, A., Drapeau, V., Bouchard, C., Vohl, M.C., Pérusse, L.: GAD2 gene sequence variations are associated with eating behaviors and weight gain in women from the Quebec family study. Physiol. Behav. August 15 (Epub ahead of print) (2009)
32. Witchel, S.F., White, C., Libman, I.: Association of the -243 A–>G polymorphism of the glutamate decarboxylase 2 gene with obesity in girls with premature pubarche. Fertil. Steril. 91, 1869–1876 (2009)
33. Groves, C.J., Zeggini, E., Walker, M., Hitman, G.A., Levy, J.C., O'Rahilly, S., Hattersley, A.T., McCarthy, M.I., Wiltshire, S.: Significant linkage of BMI to chromosome 10p in the U.K. population and evaluation of GAD2 as a positional candidate. Diabetes 55, 1884–1889 (2006)

A Novel Suppression Operator Used in optaiNet

Jungan Chen, Feng Liang, and Wenxin Chen

Electronic Information Department, Zhejiang Wanli University
No.8 South Qian Hu Road, Ningbo, Zhejiang, 315100, China
{friendcen21,liangf_hz}@hotmail.com, yulong@zwu.edu.cn

Abstract. optaiNet is proposed to function optimization. A threshold is used to control the network cells suppression in optaiNet. But the threshold is required to set manually by experience. In this paper, a novel suppression operator is proposed and used to make an improvement in optaiNet. So there is no threshold in the improved algorithm. The comparison experiment is conducted. The results show that the novel suppression operator is valid. The improved algorithm can achieve the optimized network size and is more effective than optaiNet.

Keywords: artificial immune system; function optimization; negative selection; clonal selection.

1 Introduction

Nowadays, there are a lot of applications of artificial immune system(AIS) including classification, optimization, computer security and learning. Negative selection, clonal selection, immune network, danger theory are the main principles presented in AIS.

Clonal selection algorithm (CSA) is first proposed by de Castro and proved to be capable of solving complex machine learning tasks, like pattern recognition and multimodal optimization [1]. Clone and selection are the main operators in CSA. Clone operator includes three mechanisms which are clonal expansion (producing identified cells with its parent), somatic hypermutation and receptor editing. Somatic mutation guides to local optima, while receptor editing introduce diversity. Selection operator is used to form the memory cell and maintain global or local optima. Later, the clonal selection algorithm is named CLONALG. By comparing with a niching method for multimodal function optimization, it shows that the CLONALG can reach a diverse set of local optima solutions [2].

By combining clonal selection with immune network, aiNet is proposed to clustering [3]. For the implementation of an optimization version of aiNet, optaiNet is proposed[4]. Comparing optaiNet with CLONALG, optaiNet performs a better exploration of the search space. But it is metioned that “As future trends for the optimization version of aiNet, several aspects still have to be accounted for. First, the algorithm sensitivity to its tuning parameters must be assessed [4]”.

In optaiNet, clone and selection operator is used to network expansion; network suppression is controlled by suppression threshold which should be set manual by experience. In this paper, a novel suppression operator is proposed and used to make an

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 17 – 23, 2009.

improvement of optaiNet. The suppression operator is called Negative Selection operator (NS). In human immune system, through negative selection, T-cells which recognize self cells are destroyed before leaving the thymus. Inspired from this, Negative selection algorithm (NSA) is proposed and mainly used to anomaly detection [5]. In this work, NSA is a part of optaiNet applied to function optimization. NS operator is used to eliminate those network cells which are recognized by others. So there is no suppression threshold in the new algorithm called optaiNet with negative selection operator (optaiNet-NS).

2 The Model of the Algorithm

In optaiNet-NS, population composed of network cells is encoded with real values. Fitness is calculated through the objective function. Affinity is the Euclidean distance between two network cells. Clone means the copy of its parent. The value r is the match threshold of one network cell.

Recognition is defined as following:

If Affinity(cell1,cell2)<cell1.r, cell1 recognizes cell2
If Affinity(cell1,cell2)<cell2.r, cell2 recognizes cell1

NS operator is defined as following:

When recognition happens, the cell with lower fitness is removed from the population.

The model of optaiNet-NS is following,:

```
1. Randomly initialize a population of cells CELLS
2. While(true)
2.1 calculate the fitness of each cell CELLS[i] and the
average fitness AVGFIT of CELLS.
2.2 //network cells expansion
2.2.1 For each cell CELLS[i]
//clonal operator
2.2.2 Generate a number Nc of clones CLONES[i].
2.2.3 Mutate each new cells CLONES[i][j] proportionally to
the fitness of its parent cell. The mutation follows
equation (1).
2.2.4 Calculate the fitness of all new cells in
CLONES[i][j].
// selection operator
2.2.5 Select the cell with highest fitness
CLONES[i][highest] within CLONES[i].
2.2.6 Set CLONES [i][ highest].r=min(CLONES [i][
highest].r, affinity(CLONES[i][ highest], CELLS [i])).
2.2.7 Compare CLONES [i][ highest] with CELLS[i], add the
cell with highest fitness to new population NEWS
2.2.8 EndFor
2.3 Set CELLS = NEWS
2.4 Calculate the average fitness AVGFITnew of CELLS.
```

```
2.5 If AVGFITnew-AVGFIT>0.0001 , AVGFIT= AVGFITnew, go to
2.2.1
2.6 //network cell suppression(NS operator)
2.6.1 For each cell CELLS[i] and others CELLS[j] ,
j=[1…,i-1,,i+1,…]
2.6.2 If affinity(CELLS[i],CELLS[*])<CELLS[i].r or
affinity (CELLS[i],CELLS[*])<CELLS[*].r, remove the cell
with lowest fitness from CELLS
2.6.3 EndFor
2.7 Introduce a percentage d% of randomly generated cells
and
return to step 2.
2.8 If stopping criterion is met, break;
3. EndWhile
```

In step 1, the initial population consists of N network cells. Then network begins to expansion in step 2.2. During steps 2.2.1~2.2.3, each network cell undergoes a process of clonal operator which includes clonal expansion and affinity maturation. In affinity maturation, clones of each cell are mutated according to the affinity of the parent cell. The fitness represents the value of the function for the specific candidate solution. The affinity proportion mutation is performed according to the following equation:

$$C^{'} = c + \alpha N(0.1) \tag{1}$$

$$\alpha = (1/\beta)\exp(-f^{*}) \tag{2}$$

where α, is the amount of mutation, c is the parent cell, c' is the mutated clone of c, N(0,1) is a Gaussian random variable of zero mean and standard deviation of 1, βis parameter that controls the decay of the inverse exponential function and f* is the fitness of c normalized in the interval [0..1]. As c' represents a candidate solution, it must be within the range of the functions specified domain. If c' exceeds that, then it is rejected and removed from the population [6].

During steps 2.2.4~2.2.8, The fitness of each new cells produce by clonal operator is evaluated, then the cell with highest fitness being selected to become a memory cell, which is called selection operator.

Steps 2.3~2.5 are used to lead the network expansion to a stabilization. It is known that the average fitness can reflect the stabilization situation of the population. When the degree of similarity between the present average fitness (AVGFITnew) and the previous average fitness(AVGFITNESS) is less than 0.0001, the network is taken as stable[4]. Once stabilization occurs, network suppression in steps 2.6.1~2.6.3, where Negative Selection operator (NS) is used, is processed. In NS operator, if any network cell CELLS[i] is recognized by other cell CELLS[j] according the definition of recognition before, then the cell with lower fitness between CELLS[i] and CELLS[j] is remove from the network. Step 2.7 is used to receptor editing which is part of clonal operator.

3 Experiments

The objective of the experiments is to:

1. Investigate whether NS operator is valid
2. Compares optaiNet-NS with optaiNet

$$a = \begin{bmatrix} -32 & -16 & 0 & 16 & 32 & -32 & -16 & \dots & 0 & 16 & 32 \\ -32 & -32 & -32 & -32 & -32 & -16 & -16 & \dots & 32 & 32 & 32 \end{bmatrix} \tag{3}$$

$$f = 0.002 + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^{2} (x_i - a_{ij})^6}, x_i \in [-65.356, 65.356] \tag{4}$$

For comparison, shekel's function shown in the equation (4) is used to optimize. Algorithms run 10 times. The parameters are set as following: N = 10; Nc = 10; d= 0.1; β= 1. In optaiNet, supressionThreshold is set to 0.0005, 0.001, 0.01, 0.1, 0.15, 0.2 respectly. The results are shown in fig.1~5.

It is known that there are 25 minimal optimal values in shekel's function. The optimal values found by optaiNet-NS are shown in fig.1, so the NS operator is valid. The optimal values found by optaiNet with different threshold are shown in fig.2. Less than 15 optimal values are found when supressionThreshold is set to 0.1,0.15,0.2. So optaiNet is sensitivity to the supressionThreshold.

The average sizes of the population (i.e. optimal values) produced by both algorithms are shown in fig.3 and the average fitness are shown in fig.4. OptaiNet-NS finds about 27 optimal values in fig.3, With different threshold, optaiNet finds different quantity optimal values. When supressionThreshold is set to 0.1,0.15,0.2, less quantity optimal values are found. Otherwise, More than 28 optimal values are found. So the

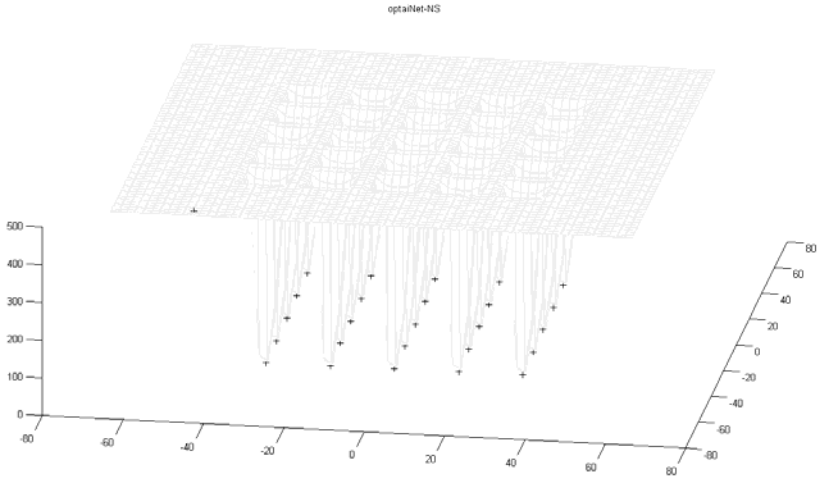


Fig. 1. Optimal value found by optaiNet-NS

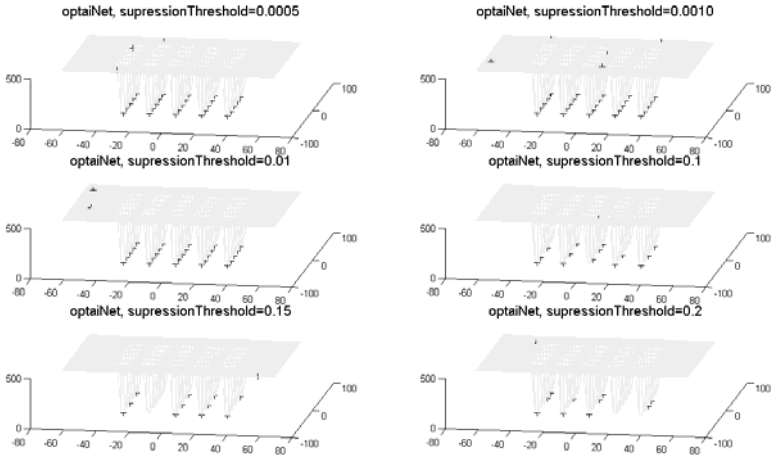


Fig. 2. Optimal value found by optaiNet

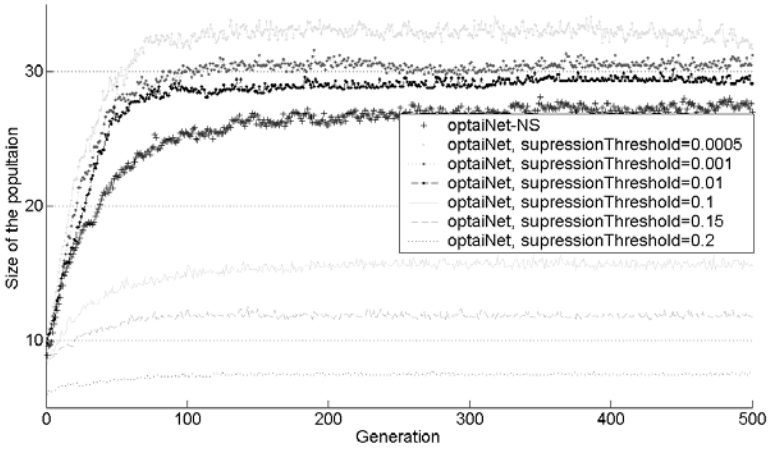


Fig. 3. Size of the population

results in fig.2,3 show that optaiNet is sensitivity to the value of supressionThreshold. Because the value of supressionThreshold is set by manually, optaiNet is difficult to be applied. From fig.1 ~fig.4, it is concluded that optaiNet-NS can achieve the optimized network size in fig.3 and find all the optimal value in fig.1.

With optaiNet-NS, the minimal value of r is converged to a stable value in fig.5. It leads the population converge to the different optimal value.

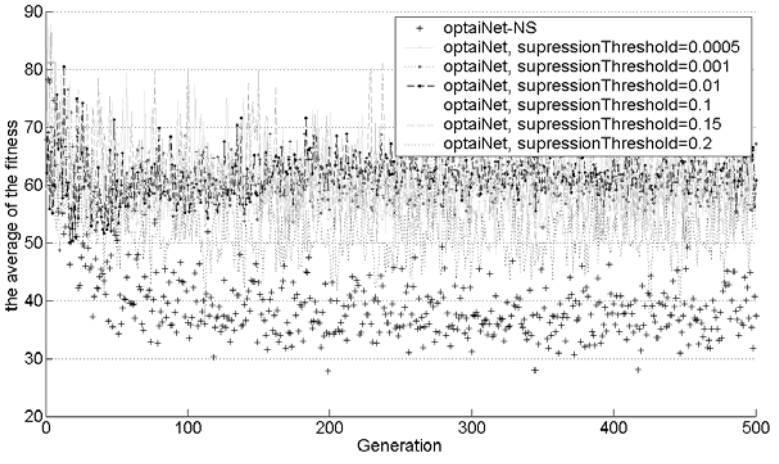


Fig. 4. The result of the average fitness

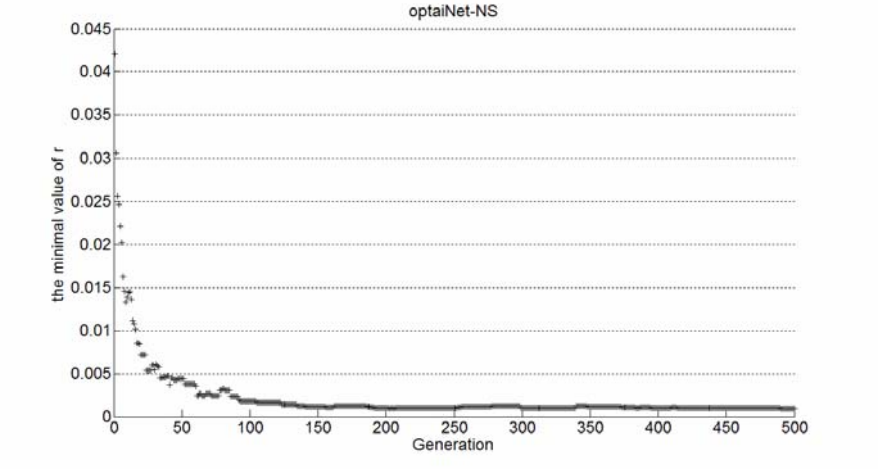


Fig. 5. The result of the minimal value of r

4 Conclusion

In this paper, a novel suppression operator called negative selection operator(NS) is proposed and used to make an improvement of optaiNet. So there is no threshold in the improved algorithm optiaiNet-NS. The results of experiment show that NS operator is valid. As optaiNet-NS can achieve the optimized network size and find all the optimal value, it is more effective than optaiNet.

Of course, a lot of algorithms about clonal selection algorithm have been proposed[7], so it is necessary to compare optaiNet-NS with other algorithms in the future.

Acknowledgement

This work is supported by Ningbo Nature Science Foundation 200701A6301043 and Scientific Research Fund of Zhejiang Provincial Education Department 20070731. Thanks for the source code supplied by Jason Brownlee [http://www.artificial-immune-systems.org/algorithms.shtml].

References

[1] de Castro, L.N., Von Zuben, F.J.: The clonal selection algorithm with engineering applications. In: Proceedings of Genetic and Evolutionary Computation Conference, Las Vegas, USA, pp. 36–42 (2000)

[2] de Castro, L.N., Von Zuben, F.J.: Learning and optimization using the clonal selection principle. IEEE Transaction on Evolutionary Computation 6(3), 239–251 (2002)

[3] de Castro, L.N., Von Zuben, F.J.: aiNet: An Artificial Immune Network for Data Analysis, p. 6. Idea Group Publishing, USA (2001)

[4] de Castro, L.N., Timmis, J.: An artificial immune network for multimodal function optimization. In: Proceedings of the 2002 Congress on Evolutionary Computation, pp. 699–704 (2002)

[5] Forrest, S., Perelson, A.S., Allen, L., Cherukuri, R.: Self-nonself Discrimination in a Computer. In: Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society, Los Alamitos (1994)

[6] Timmis, J., Edmonds, C.: A Comment on opt-AiNET: An Immune Network Algorithm for Optimisation. In: Proc. of the CEC conf., San Diego, pp. 308–317 (2004)

[7] Brownlee, J.: Clonal Selection Algorithms [Technical Report]. Victoria, Australia: Complex Intelligent Systems Laboratory (CIS), Centre for Information Technology Research (CITR), Faculty of Information and Communication Technologies (ICT), Swinburne University of Technology (2007)

Principal Subspace Analysis Based BCG Artifact Removal in Single Channel EEG Signal Measured Inside MRI Scanner

Tahir Rasheed*, Young-Koo Lee, and Sungyoung Lee

Department of computer Engineering
Kyung Hee University, Suwon, 449-701, Repubic of Korea
{tahir,yklee,sylee}@oslab.khu.ac.kr

Abstract. Single channel EEG analysis is vital for clinical as well as for brain computer interface (BCI) studies. The measured EEG signal contains different artifacts. Among these artifacts Ballistocardiogram artifact is most prominent; it gets amplified when measurements are made inside the MRI scanner making the EEG analysis practically impossible. There are different methods to remove these artifacts from single channel observation. However, these conventional methods either require an estimation of artifact template or reference signal for the artifact. In this study, we propose a method based on principal subspace analysis for BCG artifact removal from EEG signal measured inside MRI. This method does not suffer from any of the above mentioned disadvantages of conventional methods. We have removed the BCG artifacts from both the continuous as well as visual evoked potentials measured inside MRI scanner. The results presented in the manuscript suggest that the proposed method could be used for single channel EEG studies.

Keywords: Electroencephalography (EEG), Principal Subspace Analysis (PSA), Artifact Removal, Ballistocardiogram (BCG).

1 Introduction

Electromagnetic (EM) brain signals represent the functionality of underlying sources. By combining the two modalities, electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), brain activities can be mapped at superior spatiotemporal resolution. However, one of the limitations is that EEG signal measured inside the MR scanner get significantly corrupted by artifacts: most significant of which are gradient, ballistocardiogram (BCG) and electro-oculogram (EOG) artifacts. It is known that the gradient artifact is due to changing fMRI magnetic fields, the BCG artifact due to the tiny movement of EEG electrodes inside the MRI scanner because of the pulsatile changes in blood flow tied to cardiac cycle and the EOG artifact by the eyes movement of the subject. It has been reported that the magnitude of these artifacts is much higher compared with the alpha rhythm of EEG [1] [2].

* Corresponding Author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 24–30, 2009.

Allen et al in 1998 [1], was the fist who proposed a method to remove BCG artifacts. In this method, know as average artifact subtraction method (AAS), an artifact template was obtained by averaging the artifacts per heart beat and then subtracting from the EEG signal. The standard AAS method is the most common technique used in available commercial software for artifact removal. Attempts have been made using this procedure not only for spontaneous EEG but also to recover the visual evoked potentials measured during fMRI [3] [4]. One critical requirement for AAS is the simultaneous acquisition of ECG to identify each heart beat. However, as the ECG is a non-stationary signal and is affected by the magnetic field as well, this method is associated with less representative templates. Attempts for real-time artifact removal have also been made. In one such attempt, motion sensors were used to measure the head movements and the adaptive filters were utilized to remove the artifacts [2]. Later Kalman filtering of EEG signals, using EOG channel as reference, for the BCG artifact removal has been done by our group [5]. Usually these adaptive filtering techniques assume known variances and therefore require reference channels for generating the artifacts. A statistical method, independent component analysis have also been applied for artifact removal from multichannel observed EEG signals [6] [7]. However, little work is available for single channel artifact removal using ICA analysis. The interest in sparse signal representation is growing, in which the signals are decomposed into several sparse components. Recently in 2009, Yong *et. al.* [8] presented a technique based on sparse signal representation for artifact removal form single channel EEG.

In comparison to biomedical signals sufficient amount of work is available for single channel audio signal extraction. In 2000, Casey and his colleague introduced independent subspace analysis (ISA) [9] for separation of mixed audio sources. Motivated by ISA; in this study, we propose a method based on principal subspace analysis (PSA) for single channel artifact removal for EEG signal measure inside MRI. The proposed method has a advantage over the conventional single channel techniques; it does not require templates of artifacts or reference signals. Based on the results we believe that the proposed artifact removal method could be an effective tool for single channel EEG studies measured inside MRI.

2 Experimental Setup

Continuous as well as Visual evoked potentials upon checker-board reversals (1 or 2Hz) were recorded from four healthy volunteers inside and outside the 3.0T whole body MRI scanner (Magnum 3.0, Medinus, Korea). The volunteers (four men, mean age of 26.6) with no previous history of neurological and psychiatric disturbance were recruited from an academic environment. We used a MRI-compatible 32-channel EEG recording system (BrainAmp MR, Brain Products GmbH, Germany) for EEG data acquisition. The EEG signal was amplified and then transformed into optical signal in the EEG amplifier to be transmitted to the EEG data acquisition system placed outside the MRI shield room. We used the sampling rate

of 1 KHz and the bandwidth of 1-60Hz for band-pass filtering. All the EEG recordings were performed with the standard 10-20 uni-polar system referenced to FCz. Electrode skin impedance was kept below 1KHz. To minimize motion artifact in EEG on the scalp electrode of the subject, we tightly fixed the EEG cap on the scalp using adhesive tapes. Furthermore, to minimize the motion artifacts of the EEG lead wires between the EEG cap and the EEG amplifier, we fixed the lead wires to a supportive structure using plastic ties. The study was approved by the institutional ethics review committee of Kyung Hee University, Korea, and written informed consent was obtained from each subject.

3 Principal Subspace Analysis

The principal subspace analysis operates on a assumption that the source mixture signal is composed of n principal sources,

$$c(t) = \sum_{j=1}^{n} c_j(t) \tag{1}$$

The input signal *c(t)* is split into finite-length segments; $c^{(k)}$, where $1 \leq k \leq m$ is the segment index. Each segment is multiplied by a transformation matrix M. The absolute value of the transformed segment k produces an observation vector $\mathbf{x} \in \mathbf{R^n}$ and all the frames makes a complete spectrogram. The transformation used is the fourier transform.

$$\mathbf{x}^{(k)} = M^t \mathbf{c}^{(k)} \tag{2}$$

The column of a spectrogram corresponds to spectral slice which is a snapshot of the spectrum at time k. Each frame of the input spectrogram can be expressed as a weighted sum of ρ principle basis vectors, $\mathbf{e}_i$. The basis vectors remains constant but their weighted sum reconstructs a spectrogram frame:

$$\mathbf{x}^{(k)} = \sum_{i=1}^{\rho} \mathbf{y}_i^{(k)} \mathbf{e}_i \tag{3}$$

The principal basis are orthonormal, the weight coefficients $\mathbf{Y}$ for all the frames of the spectrogram can be obtained by projecting the input (spectrogram) onto basis components:

$$\mathbf{Y} = \mathbf{E}^t \mathbf{X} \tag{4}$$

The input spectrogram can be decomposed into sums of principal subspaces according to the following equation:

$$\mathbf{X} = \mathbf{e}_1 {\mathbf{y}_1}^t + \mathbf{e}_2 {\mathbf{y}_2}^t + \ldots + \mathbf{e}_n {\mathbf{y}_n}^t \tag{5}$$

Each spectrogram $\mathbf{X}_j = \mathbf{e}_j \mathbf{y}_j^t$ corresponding to a subspace, obtained from a set of basis vectors, is transformed back by inverting the transformation carried out by Equation 2.

$$\mathbf{s}_j = M^{-1} \mathbf{X}_j \tag{6}$$

By all the above mentioned procedure we have divided the single mixture into number of components that corresponds to principal subspace in the transformed domain. Grouping of these time domain components is carried out.

$$\mathbf{S}_j = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_p], \;\; \mathbf{S}_j \subseteq \{\mathbf{S}\} \tag{7}$$

Finally, the separated source signal are obtained by point to point sum of the signals in each group.

$$\mathbf{c}_i = \sum_{j=1}^{p} \mathbf{s}_j \tag{8}$$

The grouping criteria depends on the application area. In our implementation we have grouped the components into two groups; depending on the ordering of the components.

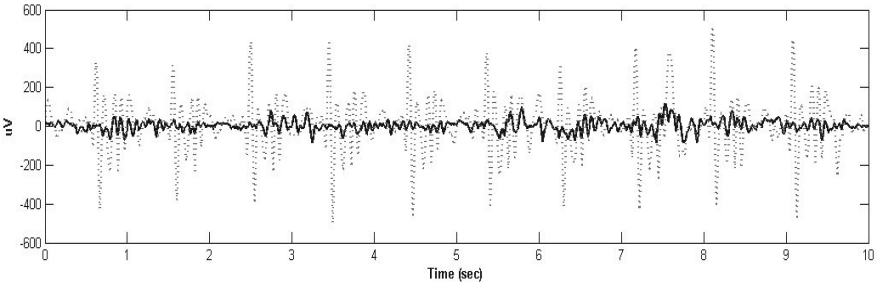
4 Results

Single channel EEG signal is segmented and transformed into the frequency domain using the fourier transform. The segments should be short enough so that the segmented signal is stationary. The magnitude and angles are calculated. The magnitude forms the spectrogram that needs to be split and the angle information is used during the inverse fourier transform. Depending on the length of the input EEG signal, data reduction may also be required. In our study we did the data reduction by keeping the 80% of the variance. The input EEG signal is decomposed in to source signals according to the procedure mentioned in the section 3. As the principal subspaces in the spectrogram domain are ordered according to variance, i.e., high variance to low variance or vice versa,that is way the source signals in time domain are also ordered accordingly.

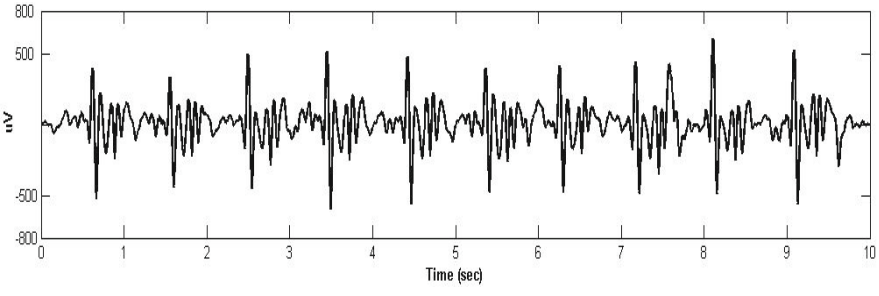
The BCG artifacts are due to pulsatile changes in blood flow tied to the cardiac cycle inside the MRI scanner. These artifacts have high amplitudes (variance) as compared to the normal EEG. Therefore, the first two to three source signals (i.e., components that corresponds to high variance) can capture these artifacts and be placed into the first group. The rest all components are placed into the second group.

4.1 Continuous EEG

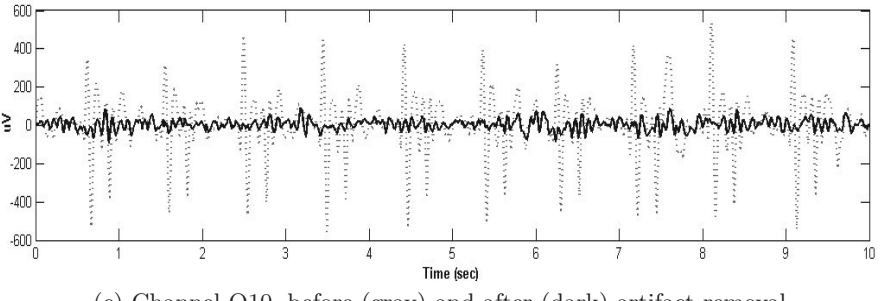
Figure 1, shows the results of BCG artifact removal from two representative channels O9 and O10. Figure 1(a), shows the original channel O9 (gray) overlayed with the signal after artifact removal (dark). Figure 1(b), shows the recovered BCG artifact signal. Similarly, Figure 1(c) shows the original channel O10 (gray) overlayed with the signal after artifact removal (dark) and Figrue 1(d) shows the recovered BCG artifact. The results clearly depicts that the BCG artifact is well extracted by the proposed method.



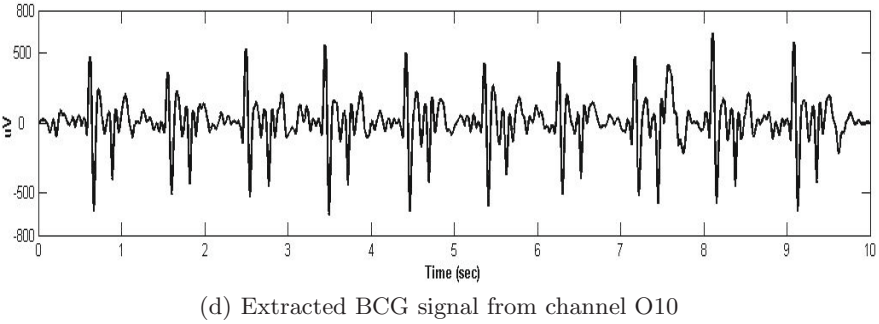
(a) Channel O9, before (gray) and after (dark) artifact removal



(b) Extracted BCG signal from channel O9

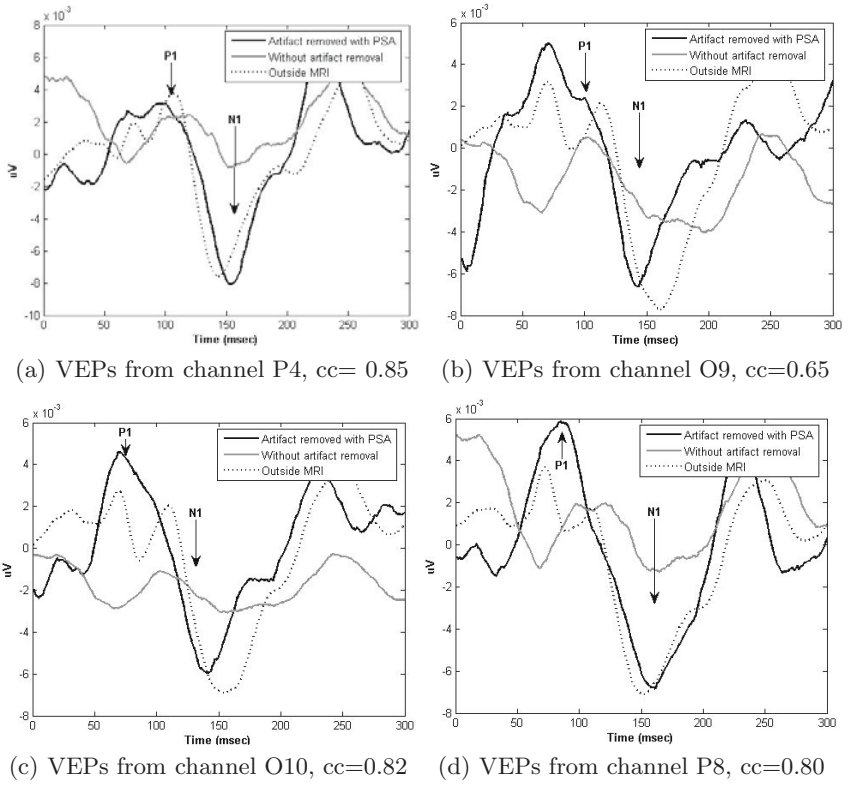


(c) Channel O10, before (gray) and after (dark) artifact removal



(d) Extracted BCG signal from channel O10

Fig. 1. Channels O9 and O10, signals before (gray) and after (dark) artifact removal and the extracted BCG signal



(a) VEPs from channel P4, cc= 0.85 (b) VEPs from channel O9, cc=0.65

(c) VEPs from channel O10, cc=0.82 (d) VEPs from channel P8, cc=0.80

Fig. 2. Comparison between VEPs obtained after artifact removal (solid), without artifact removal (gray) and outside MRI (dotted)

4.2 Visual Evoked EEG

In the case of evoked potentials, EEG signal was averaged according to the event timings after artifact removal. The recovered VEP is compared against the VEPs from the EEG signal acquired outside MRI (taken as the gold standard) under the identical experimental settings. The VEPs from some representative channels, before and after artifact removal against the outside VEPs are shown in Figure 2. The results clearly shows that the VEPs obtained after artifact removal are much similar to those of outside MRI VEPs. The computed cc (correlation coefficient) values indicate a high similarity between the after artifact removed VEPs and outside MRI VEPs. It is clear that the recovered VEPs are much similar to the outside VEPs. The P1ŰN1 complex is also detected in all cases.

5 Discussion and Conclusions

In this study, single channel artifact removal method is presented. The proposed method decomposes a mixture signal into a number of signals based on

principal subspace analysis. Grouping these signals according to some criteria separate the mixture signal into constituents. The proposed method does not require any template or reference signals thus overcoming the already described disadvantages posed by conventional methods. In this study we presented the idea, in future we plan to compare its performance with already existing methods and to extend this method for EEG measured outside MRI scanner. Based on the results we believe that the proposed methods may be an effective tool for EEG analysis.

Acknowlegment

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement) (IITA-2009-(C1090-0902-0002)). This work also, was supported by the Korea Science & Engineering Foundation(KOSEF) grant funded by the Korea government (MEST)(No. 2008-1342), and was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0076798).

References

1. Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L.: Identification of EEG events in the MR scanner: The problem of pulseartifact and a method for its subtraction. NeuroImage 8, 229–239 (1998)
2. Bonmassar, G., Purdon, P.L., Jaaskelainen, I.P., Chiappa, K., Solo, V., Brown, E.N., Belliveau, J.W.: Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI. Neuroimage 16, 1127–1141 (2002)
3. Comi, E., Annovazz, P., Silva, A.M., Cursi, M., Blasi, V., Cadioli, M., Inuggi, A., Falini, A., Comi, G., Leocani, L.: Visual evoked potentials be recorded simultaneously with fMRI scanning: A validation study. Human Brain Mapping 24, 291–298 (2005)
4. Becker, R., Ritter, P., Moosmann, M., Villringer, A.: Visual evoked potentials recovered from fMRI scan periods. Human Brain Mapping 26, 221–230 (2005)
5. In, M.H., Lee, S.Y., Park, T.S., Kim, T.S., Cho, M.H.: Ballistocardiogram artifact removal from EEG signals using adaptive filtering of EOG signals. Physiological measurements 27, 1227–1240 (2006)
6. Srivastava, G., Crottaz-Herette, S., Lau, K.M., Glover, G.H., Menon, V.: ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner. Neuroimage 24, 50–60 (2005)
7. Mantini, D., Perrucci, M.G., Cugini, S., Ferreti, A., Romani, G.L., Gratta, C.D.: Complete artifact removal for EEG recorded during continuos fMRI using independent component analysis. NeuroImage 31, 598–607 (2007)
8. Yong, X., Ward, R.K., Birch, G.E.: Artifact removal in EEG using morphological component analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing, April 2009, pp. 345–348 (2009)
9. Casey, M.A., Westner, A.: Separation of mixed audio sources by independent subspace analysis. In: Proceedings of the International Computer Music Conference, Berlin (August 2000)

Genetical Analysis of Ascochyta Blight Resistance in Chickpea

Alireza Taleei[1,*], Homayoun Kanouni[2], and Michael Baum[3]

[1] Professor in the Department of Agronomy & Plant breeding,
Faculty of Crop & Animal Sciences, College of Agriculture and Natural Resources,
University of Tehran, Karaj, P.O. Box 31587-11167, Iran
ataleei@ut.ac.ir
[2] Seed and Plant Improvement Institute (SPII), 4119-31585, Karaj, Iran
[3] International Centre for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria

Abstract. Ascochyta blight, caused by *Ascochyta rabiei* (Pass.) Lab. is a devastating disease of chickpea (*Cicer arietinum* L.) worldwide. Available genetic variation for Ascochyta blight resistance in genus *Cicer* has prompted interest in the development and use of resistant cultivars that can be sown in autumn and, to increase seed yield in chickpea. Understanding the mode of inheritance of resistance to Ascochyta blight in chickpea would assist breeding efforts. The objective of this study were determining number of genes confer Ascochyta blight resistance and leaf size as well as action of them. Thus F_1, F_2 F_3 progenies derived from a cross between Iranian local variety Bivanij (susceptible local variety) and ICC12004 along with their parents were sown in a RCB design at the International Center for Agricultural Research at Dry Area (ICARDA) under artificial infection conditions. Results showed that in F_2 and F_2 generations the ratio of susceptibility did not differ significantly from those of 9:7 and 5:3 theoretical ratios. There was a negative correlation between leaf size and blight score, which means that large leaf genotypes could be more susceptible to Ascochyta blight. Generation mean analysis for resistance to Ascochyta blight in this study revealed that additive effect has main role in Ascochyta blight resistant, however the leaf size besides of additive effect showed dominance effect as well. For theses traits we detected dominant x dominant interaction (l) in the opposite sign which reveals the evidence of a duplicate epistasis. These findings showed that the genotype of resistant parent could be as $R_1R_1R_2R_2$. According to these findings and available knowledge, it would be suggested an appropriate breeding program for gene pyramiding to produce multiple resistant genotypes in chickpea.

Keywords: Chickpea (*Cicer arietinum* L.), *Ascochyta rabiei* (Pass.) Lab., disease resistance, SSR, linkage map, QTL.

1 Introduction

Chickpea (*Cicer arietinum* L.), a self-pollinating diploid annual, with 2x=2n=16 chromosomes. It is the third most important grain legume in the world after common

* Corresponding author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 31–37, 2009.

bean (*Phaseolus vulgaris* L.) and pea (*Pisum sativum* L.)[19]. Primarily, chickpeas are grown in the Indian subcontinent, West Asia, North Africa, Ethiopia, Southern Europe, Mexico, Australia, North-Western United States and in the Brown and Dark Brown soil zones on the Canadian parties [8]. Average yield of chickpeas worldwide is about 700 kg/ha which is much below its potential [8], [16]. Yields are seen as low and unstable compared to other crops due to adverse effects of a number of biotic and abiotic stresses [8]. One of the greatest biotic stresses reducing potential yield in chickpea is ascochyta blight, caused by the fungus *Didymella rabiei* (Kovachevski) v. Arx. (anamorph: *Ascochyta rabiei* (Pass.) Labrousse) is the most devastating worldwide, causing up to 100 per cent yield losses in severely affected fields [7]. *Ascochyta rabiei* is heterothallic, thus when two compatible mating types are present genetic recombination can occur resulting in ascospore production [20], [21], [27]. Isolates of both mating types found in Iran indicating the occurrence of sexual recombination. Recombination could potentially lead to greater genetic and pathogenic variability in populations of *A. rabiei*. Pathogenic variability in *A. rabiei* populations has been reported in almost all chickpea growing regions in the world, including India, Iran, Pakistan, Turkey, Syria, the Palouse region of north-western United States and Canada [3], [6], [8], [12], [23]. Chongo et al. (2004) also confirmed the presence of genetic variability among *A. rabiei* isolates collected in the 1998 and 1999 growing seasons based on RAPD molecular markers [3]. Despite recognition of destructiveness of *A. rabiei* in chickpea production world-wide, very little head way on controlling the disease through resistance breeding has been made in the past century. Resistance in breeding lines of chickpea to ascochyta blight is not durable due to the high variability of *A. rabiei* populations wherever chickpeas are grown [8], [12], [14], [15], [23]. Resistance break down is possibly the greatest challenge in breeding for resistance to ascochyta blight in chickpea [13]. Cultivars available in ICARDA, lack complete resistance to *A. rabiei*. Partial resistance in cultivars adapted to the western Iran tends to break down after the onset of flowering. Partially resistant cultivars contribute to the development of new pathotypes of the disease by imposing selection pressure, possibly resulting in increased virulence or aggressiveness within the pathogen population [17]. With a genetically diverse population of *A. rabiei*, it is important not only to develop cultivars with durable forms of resistance, but also to monitor changes in the population structure to anticipate resistance breakdown in existing cultivars. Among current understanding of the genetics of ascochyta blight resistance (ABR) in chickpea strongly suggests polygenic inheritance of the trait. In an interspecific genetic background, Santra et al. (2000) mapped two QTLs which conditioned ABR over two years of field screening [9]. Likewise, preliminary QTL mapping in a wide-cross between *C. arietinum* and *Cicer echinospermum* (resistance source) revealed two to three QTLs for seedling resistance in controlled glasshouse bioassays [14]. Tar'an et al. (2007) identified one QTL on each of LG3, LG4 and LG6 accounted for 13%, 29% and 12% respectively, of the total estimated phenotypic variation for the reaction to ascochyta blight [18]. Although the genetic mechanism of ABR has been studied in identified resistant accessions of *C. arietinum*, the number

and genomic locations of the genes or QTLs conditioning resistance has yet to be verified. The objective of this study were determining number of genes confer Ascochyta blight resistance and leaf size as well as action of them. Thus F_1, F_2, F_3 progenies derived from a cross between Iranian local variety Bivanij (susceptible local variety) and ICC12004 along with their parents were sown in a RCB design at the International Center for Agricultural Research at Dry Area (ICARDA) under artificial infection conditions.

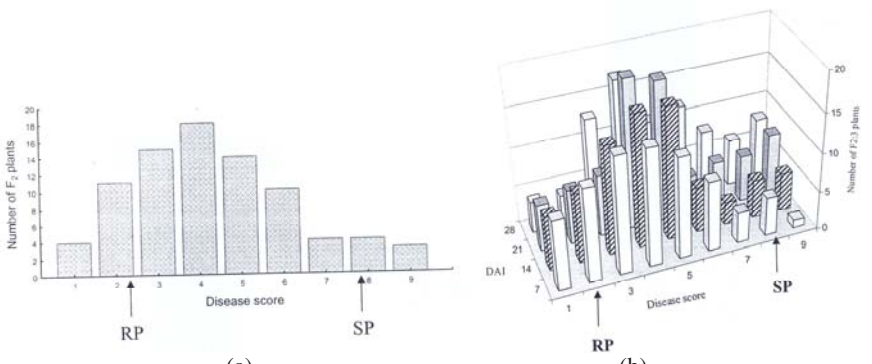
2 Material and Methods

F_1, F_2, F_3 progenies derived from a cross between Iranian local variety Bivanij (susceptible local variety) and an Indian accession ICC12004 along with their parents were sown in a Completely Randomized Block design (CRBD) at the International Center for Agricultural Research at Dry Area (ICARDA) under artificial infection conditions.

Bivanij is a high-yielding cultivar of *Kabuli* type with beige, relatively large seeds (400 mg), highly susceptible to *D. rabiei* and semi-erect growth habit. ICC12004 is resistant to the blight, with typical *Desi* small seeds (250 mg) and an erect growth habit. Isolate No. 13 of pathotype III (Udupa et al., 1998) was used for inoculation in both methods [1], [8], [25]. This isolate was cultured at room temperature under florescent light [2]. For every generation, the inoculation method was based on Buchwaldt et al., 2007, consisting in depositing a drop of spore suspension on detached leaves (10µL) [1]. Five plants of each generation were evaluated in the controlled environment. The parents, as well as the chickpea lines ILC1929 and ILC263 (susceptible), and ILC3279 (resistant to pathotypes I and II), were included as control genotypes. In this trial, the experimental design was a randomized complete block. Test plants were sown in a pair of seedling trays. Each pair of trays constituted one experimental block or replicate, and contained an individual plant of each of the $F_{2:3}$ families and control genotypes. Disease reactions were scored weekly after inoculation and AUDPC was calculated using the formula: $AUDPC=\sum [(x_i+x_{i+1})/2] (t_{i+1}- t_i)$. Isolate No.13 (PIII), was grown at room temperature under continuous fluorescent light. The suspension was filtered and adjusted to a final concentration of 2×10^5 conidia/mL using a hemacytometer.

3 Results and Discussion

Weighted analysis of variance showed that there is significantly difference for leaf size and reaction for Ascochyta blight disease (results not shown). Means along with their standard errors are tabulated in table 1. In F_2 and F_3 generations there were 109:90 and 115:81 resistant plants comparing to susceptible ones, respectively. These ratios were not significantly different from those of theoretical ones, say 9:7 and5:4, respectively. In this study according to Reddy and Singh (1993) chickpea varieties scaled as susceptible group (4/1-9/0) and resistant group (1/0-4/0), respectively.



(a) (b)

Fig. 1. The distributions of reaction to Ascochyta blight disease, for F_2 and F_3 generations (a, b, respectively). It is clear that both distributions show some skewness towards resistant parent, which is an indication of dominance for Ascochyta blight resistance controlling loci.

Table 1. Mean value of response to Ascochyta blight disease and leaf size with plant number per generations derived from crossing between 2 Chickpea inbred lines (ICC12004×Bivanij)

Generation	Plant Number	Mean ± SD		Phenotypical correlation
		Response to A. blight	Leaf size(cm2)	
P_1	40	2.25±0.018	5.39±0.21	0.12
P_2	40	7.98±0.013	12.66±0.26	0.31*
F_1	32	4.28±0.008	10.21±0.41	0.28
F_2	199	4.34±0.016	10.59±0.17	0.11
F_3	195	4.69±0.015	10.13±0.12	0.15*
F_1	32	4.28±0.008	10.21±0.41	-
RF_1	34	4.51±0.009	10.03±0.28	-

P_1 = ICC12004 (R)
P_2 = Bivanij (S)

Table 2. Estimates of gene effects for response to Ascochyta blight and leaf size in the cross between two chickpea inbred lines [ICC12004 (R) and Bivanij (S)]

Parameter	Response to A. blight	Leaf size
m	5.12±0.09	9.05±0.16
d	-2.86±0.09	-3.64±0.17
h	-2.18±0.52	5.26±0.85
i	-	-
l	1.34±0.49	4.14±0.95
X2	0.11	0.15
[h/d]	0.76	-1.45

m = mean
d = additive gene effect
h = dominance gene effect
i = additive × additive gene effect
l= dominance × dominance gene effect
[h/d] = degree of dominance

Table 3. Components of variance of Basic generations and number of effective factors (EF_1,EF_2) for response to A. blight and Leaf size in crossing between two chickpea inbred lines [ICC12004(R) , Bivanij (S)] to Ascochyta blight disease.

Parameter	Estimate	
	Response to A. blight	Leaf size
σ^2_A	4.12	6.62
σ^2_D	2.06	5.09
σ^2_E	4.05	3.67
H	0.87	0.36
Gs*	1.78	1.01
EF_1	1.64	3.33
EF_2	1.56	3.16

* k= 1.16 for selection intensity of 30%

Acknowledgment

This project was supported jointly by the University of Tehran and Agricultural Research, Education and Extension Organization (AREEO), Iran.

References

1. Buchwaldt, L., Booker, H., Gali, K.: A detached leaf assay for phenotyping of chickpea-ascochyta interaction. Agriculture and Agri-Food Canada, Saskatoon, Saskatchewan, S7N0X2 (2007)
2. Chen, W., Muehlbauer, F.J.: An Improved Technique for virulence assay of *Ascochyta rabiei* on chickpea. International chickpea and pigeonpea Newsletter 10, 31–33 (2003)
3. Chongo, G., Gossen, B.D., Buchwaldt, L., Adhikari, T., Rimmer, S.R.: Genetic diversity of *Ascochyta rabiei* in Canada. Plant Disease 88(1), 4–10 (2004)
4. Collard, B.C.Y., Pang, E.C.K., Ades, P.K., Taylor, P.W.J.: Preliminary investigation of QTLs associated with seedling resistance to ascochyta blight from *Cicer echinospermum*, a wild relative of chickpea. Theor. Appl. Genet. 107, 719–729 (2003)
5. Lincoln, S.E., Daly, M.J., Lander, E.S.: MAPMAKER/EXP version 3.0: A Tutorial and Reference Manual, 3rd edn. Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA(1993)
6. Kaiser, W.J., Muehlbauer, F.J.: Occurrence of *Ascochyta rabiei* on imported chickpeas in eastern Washington. Phytopathology 74, 1139 (1984)
7. Nene, Y.L.: A review of Ascochyta blight of chickpea (*Cicer arietinum L.)*. In: Saxena, M.C., Singh, K.B. (eds.) Ascochyta blight and winter sowing of chickpea, pp. 17–34. Martinus Nijhoff/Dr. W. Junk Publisher, The Hague (1984)
8. Reddy, M.V., Kabbabeh, S.: Pathogenic variability in *Ascochyta rabiei* (Pass.) Lab. Phytopathology Mediterranean 24, 265–266 (1984); Reddy, M.V., Singh, K.B.: Evaluation of a world collection of chickpea germplasm accessions for resistance to Ascochyta blight. Plant Dis. 68, 900–901(1985)

9. Santra, D.K., Tekeoglu, M., Ratnaparkhe, M., Kaiser, W.J., Muehlbauer, F.J.: Identification and mapping of QTLs conferring resistance to Ascochyta blight in chickpea. Crop Sci. 40, 1606–1612 (2000)
10. SAS Institute Inc.: SAS/STAT user's guide. Version 9.12. SAS Institute, Cary, North Carolina (2000)
11. Sethy, N.K., Shokeen, B., Bhatia, S.: Isolation and characterization of sequence-tagged microsatellite site markers in chickpea (*Cicer arietinum L.)*. Mol. Ecol. Notes 3, 428–430 (2003)
12. Singh, G.: Identification and designation of physiologic races of *Ascochyta rabiei* in India. Indian Phytopathology 43, 48–52 (1990)
13. Singh, K.B., Reddy, M.V.: Advances in disease resistance breeding in chickpea. Advanced Agronomy 45, 191–222 (1991)
14. Singh, K.B., Reddy, M.V.: Resistance to Six Races of *Ascochyta rabiei* in the world germplasm collection of chickpea. Crop Science 33, 186–189 (1993a)
15. Singh, K.B., Reddy, M.V.: Sources of resistance to ascochyta blight in wild Cicer species. Netherlands Journal of Plant Pathology 99, 163–167 (1993b)
16. Singh, K.B., Malhotra, R.S., Halila, M.H., Knights, E.J., Verma, M.M.: Current status and future strategy in breeding chickpea for resistance to biotic and abiotic stresses. Euphytica 73, 137–149 (1994)
17. Singh, K.B., Reddy, M.V., Haware, M.: Breeding for resistance to ascochyta blight in chickpea. In: Singh, K.B., Saxena, M.C. (eds.) Disease resistance breeding in chickpea, ICARDA, Aleppo, Syria, pp. 23–54 (1992)
18. Tar'an, B., Warkentin, T.D., Tullu, A., Vandenberg, A.: Genetic mapping of ascochyta blight resistance in chickpea (*Cicer arietinum L.)* using a simple sequence repeat linkage map. Genome 50, 26–34 (2007)
19. Tekeoglu, M., Santra, D.K., Kaiser, W.J., Muehlbauer, F.J.: Ascochyta blight resistance in three chickpea recombinant inbred line populations. Crop Sci. 40, 1251–1256 (2000)
20. Trapero-Casas, A., Kaiser, W.J.: Development of Didymella rabiei, the teleomorph of *Ascochyta rabiei*, on chickpea straw. Phytopathology 82, 1261–1266 (1992a)
21. Trapero-Casas, A., Kaiser, W.J.: Influence of temperature, wetness period, plant age, and inoculum concentration on infection and development of ascochyta blight of chickpea. Phytopathology 82, 589–596 (1992b)
22. Udupa, S.M., Baum, M.: Genetic dissection of pathotype-specific resistance to ascochyta blight disease in chickpea (*Cicer arietinum L.)* using microsatellite markers. Theor. Appl. Genet. 106, 1196–1202 (2003)
23. Vir, S., Grewal, J.S.: Physiological specialization in *Ascochyta rabiei*, the causal organism of gram blight. Indian Phytopathology 27, 265–266 (1974)
24. Wang, S., Basten, C.J., Zeng, Z.B.: Windows QTL Cartographer 2.0. Department of Statistics, North Carolina State University. Raleigh N.C. (2004), `http://statgen.ncsu.edu/qtlcart/WQTLCart.htm`
25. Udupa, S.M., Weigand, F., Saxena, M.C., Kahl, G.: Genotyping with RAPD and microsatellite markers resolves pathotype diversity in the ascochyta blight pathogen of chickpea. Theor. Appl. Genet. 97, 299–307 (1998)
26. Weising, K., Winter, P., Huttel, B., Kahl, G.: Microsatellite markers for molecular breeding. J. Crop Prod. 1, 113–143 (1998)
27. Wilson, A., Kaiser, W.: Cytology and genetics of sexual incompatibility in Didymella rabiei. Mycolgia 87, 795–804 (1995)

28. Winter, P., Pfaff, T., Udupa, S.M., Hüttel, B., Sharma, P.C., Sahi, S., Arreguin-Espinoza, R., Weigand, F., Muehlbauer, F.J., Kahl, G.: Characterization and mapping of sequence-tagged Microsatellite sites in the chickpea *(Cicer arietinum L.)* genome. Mol. Gen. Genet. (1999)
29. Winter, P., Benko-Iseppon, A.M., Huttel, B., Ratnaparkhe, M., Tullu, A., Sonnante, G., Pfaff, T., Tekeoglu, M., Santra, D., Sant, V.J., Rajesh, P.N., Kahl, G., Muehlbauer, F.J.: A linkage map of the chickpea *(Cicer arietinum L.)* genome based on the recombinant inbred lines from a C. arietinum ×C. reticulatum cross: localization of resistance genes for Fusarium races 4 and 5. Theor. Appl. Genet. 101, 1155–1163 (2000)

A PMMA Micro-suction Tool for Capsular Endoscope Using a Solid Chemical Propellant

Ho-soo Park[1,2], Kyo-in Koo[1,2], Sang-min Lee[1,2], Jae-won Ban[1,2], Hyo-young Jeong[1,3], and Dong-il Cho[1,2,3]

[1] ASRI/ISRC, Seoul National University, Seoul, Korea
[2] School of Electrical Engineering and Computer Sciences, Seoul National University, Seoul, Korea
[3] Interdisciplinary Program of Bioengineering, Seoul National University, Seoul, Korea
{phs,kkin76,sangmlee,jban,jhy0814,dicho}@snu.ac.kr

Abstract. This paper presents a PMMA (Poly (methyl methacrylate)) micro-suction tool for capsular endoscope using a solid chemical propellant, AIBN (azobisisobutyronitrile). A capsular endoscope integrated with a micro-suction tool can gather the gastro-intestinal juice for medical diagnosis. Our previous PDMS (polydimethylsiloxane) micro-suction tool has a problem of structural layer detachment between glass substrate and PDMS Venturi-tube body. In this study, PMMA is used as a material to enhance the bonding strength, which in turn results in an improved suction efficiency. The bonding strength of the PMMA and PDMS micro-suction tool are measured to be 1.6 N/m and 0.3 N/m, respectively. The value is five times larger for the PMMA micro-suction tool than that of the PDMS micro-suction tool. Also, the minimum negative pressure for suction is measured. The PMMA micro-suction tool starts to pump the diluted ink when generated negative pressure is 0.11 kPa. The value is approximately three times lower than that of the PDMS micro-suction tool. The diluted ink of 1.57 μL is successfully pumped into the PMMA micro-suction tool reservoir.

Keywords: PMMA micro-suction tool, Venturi-tube, Capsular endoscope, AIBN (Azobisisobutyronitrile).

1 Introduction

Recently, there has been increased research in developing versatile capsular endoscope to replace conventional endoscope [1, 2, 3]. Fig. 1 shows the conceptual schematic of the capsular endoscope integrated with a proposed micro-suction tool. First, the capsule endoscope is approached to the gastro-intestinal tract. Subsequently, camera module searches for the gastro-intestinal juice and the micro-suction tool gathers it.

Our research group reported a PDMS (polydimethylsiloxane) micro-suction tool for capsular endoscope in 2007 [4]. However previous PDMS micro-suction tool has a problem of structural layer detachment between glass substrate and PDMS Venturi-tube

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 38 – 44, 2009.

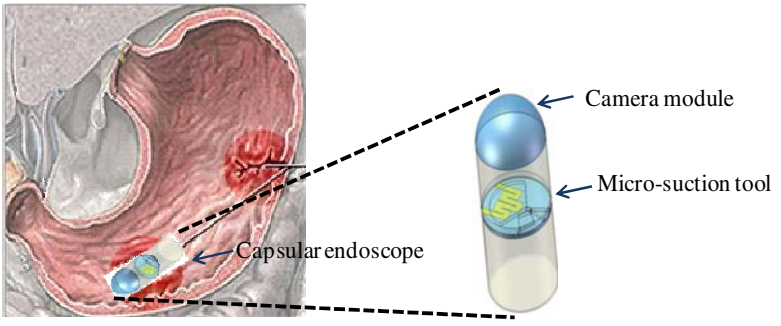


Fig. 1. Conceptual schematic of the integrated capsular endoscope

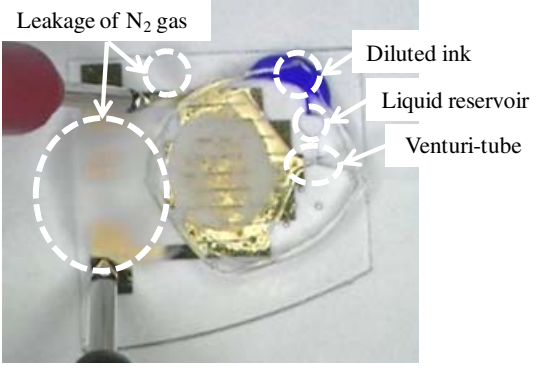


Fig. 2. N_2 gas leakage during suction

body, as shown in Fig. 2. Approximately 40 % of fabricated micro-suction tool is detached unconsciously because of the simultaneous heat and pressure generated during the suction. Therefore, the generated N_2 gas leaks to unexpected direction. Additionally, PDMS patterning is not as accurate as modern lithography tools allow, resulting in somewhat loose dimensional control, which in turn results in performance degradations.

In this paper, an advanced micro-suction tool fabricated with PMMA (Poly (methyl methacrylate)) for capsular endoscope is presented. The MMA (methyl methacrylate) is used to obtain the strong bonding strength between glass substrate and PMMA Venturi-tube body. In order to improve dimensional tolerances, we used X-ray synchrotron LiGA (Lithographie, Galvanoformung, Abformung; a German abbreviation of lithography, electroplating and molding) micromachining for the PMMA Venturi-tube body. Additionally, the contact angle between PMMA and diluted ink is 6 ° less than that of the PDMS. Therefore, the PMMA micro-suction tool is able to operate in less negative pressure than the PDMS micro-suction tool.

2 Design

A thin film of Ti/Au micro-heater is designed and fabricated for the AIBN (Azobisisobutyronitrile) ignition. AIBN matrix generates biologically-inert N_2 gas at relatively low temperature of 70 °C [5]. The decomposition formula of AIBN is:

$$C_8H_{12}N_4 \xrightarrow[heat\ over\ 70°C]{} 2C_4H_6N + N_2\uparrow \quad (1)$$

The micro-heater is designed with a length of 57.2 mm, a width of 1.4 mm, and 750 Å thick Au film. The resistance of the thin film micro-heater is designed to have 21.7 Ω, but the measured value of the resistance is 25.2 Ω. The error has occurred from the 500 Å Ti film which is an adhesion layer.

A PMMA micro-suction tool is designed utilizing a Venturi-tube design [6]. In a Venturi-tube, incompressible fluid is accelerated at the vena contracta section so that the accelerated fluid generates negative pressure. Fig. 3 shows a simple Venturi-tube channel which is originated from Bernoulli's equation:

$$\triangle P = P_1 - P_2 = \frac{1}{2}\rho v_1^2\left(\frac{A_1^2}{A_2^2} - 1\right) \quad (2)$$

Where P_1, P_2 are pressures, ρ is density, v_1 is fluid velocity, and A_1, A_2 are channel areas. To compensate for the pressure drop, suction fluid flow is generated in vena contracta section. The shape of the Venturi-tube is designed to maximize the negative pressure at the intersection of vena contracta and suction channel.

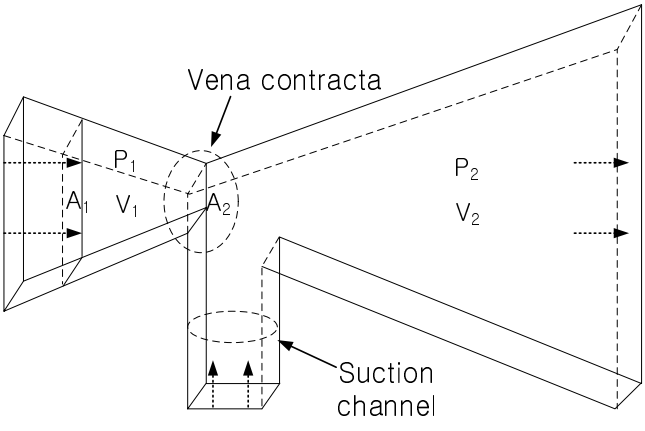


Fig. 3. Conceptual schematic of Venturi-tube channel

3 Fabrication

A PMMA micro-suction tool consists of micro-heater, AIBN chamber, Venturi-tube and liquid reservoir. After each component is fabricated, all components are bonded layer by layer with MMA. Fig. 4 shows fabrication process of the PMMA micro-suction tool. The micro-heater for the AIBN ignition is fabricated by patterning of the Ti/Au film on a 500 μm thick pyrex-7740 glass wafer [Fig. 4 (a)]. Then, liquid PMMA (950 PMMA C 9, MicroChem, America) is spin coated and baked three hours to form an adhesion layer [Fig. 4 (b)]. The AIBN powder (Unisource, India) is mixed with

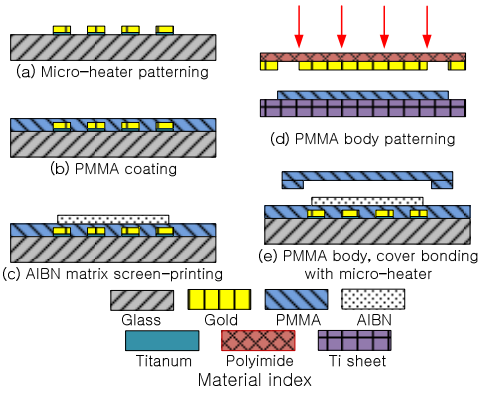


Fig. 4. Fabrication process of the PMMA micro-suction tool

teflon powder (Teflon® AF 2400, DuPont Korea, Korea) at a ratio of 5:1 in the solvent (PF-5080, 3M Korea, Korea). The AIBN matrix sol-gel is screen-printed on the micro-heater, and cured at the room temperature [Fig. 4 (c)]. The PMMA Venturi-tube body is fabricated using LiGA process [Fig. 4 (d)]. The PMMA Venturi-tube body, PMMA cover-plate and Ti/Au micro-heater are bonded using MMA interlayer [Fig. 4 (e)]. And then, approximate pressure of 100 kPa is applied for three hours for bonding.

In Fig. 5, fabrication result of the PMMA micro-suction tool is shown. The micro-heater with AIBN matrix and PMMA Venturi-tube body is shown in Fig. 5 (a) and Fig 5. (b), respectively. The bonding result of implemented PMMA micro-suction tool is shown in Fig 5. (c).

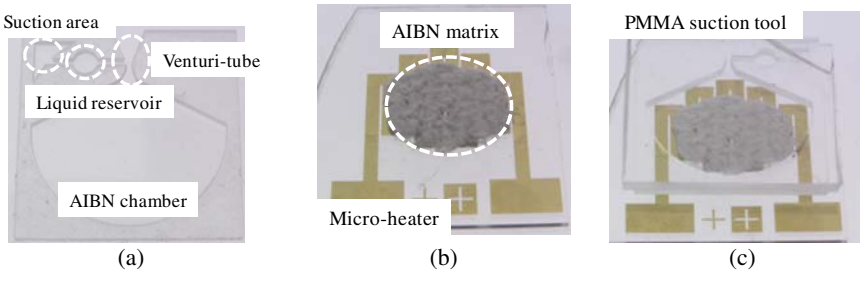
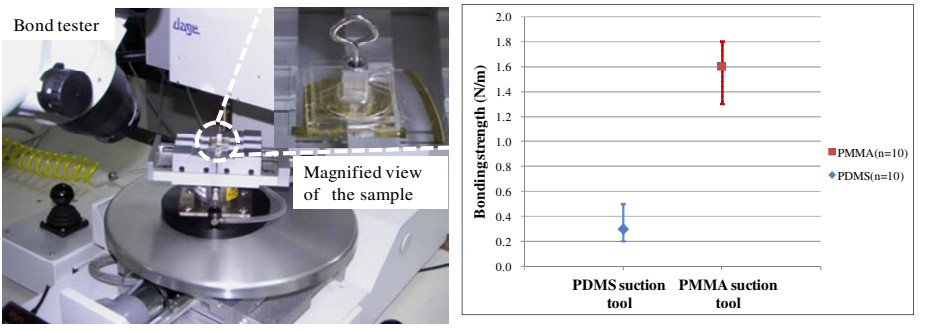


Fig. 5. Fabrication results of the PMMA micro-suction tool (a) Micro-heater with AIBN matrix (b) PMMA Venturi-tube body (c) Bonded PMMA micro-suction tool

4 Experiment

1) Measurement of bonding strength

The bonding strength of the PMMA and PDMS micro-suction tool is measured using a bond tester (DAGE 4000, DAGE, United Kingdom) [Fig. 6 (a)]. The PMMA Venturi-tube body is detached from glass substrate at 1.6 N/m and the PDMS Venturi-tube body is detached from glass substrate at 0.3 N/m, as shown in Fig. 6 (b).



(a) (b)

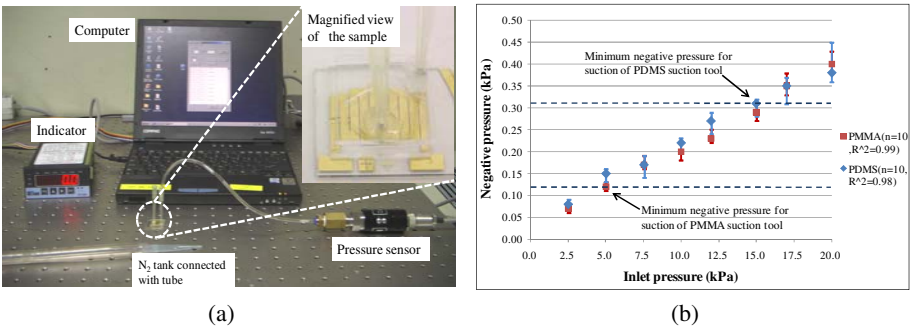
Fig. 6. Bonding strength measurement of the PMMA and PDMS micro-suction tool (a) Measurement setup (b) Measurement result of the bonding strength

The bonding strength of the PMMA micro-suction tool is about five times stronger than that of the PDMS micro-suction tool.

2) Measurement of minimum negative pressure for suction

In order to measure the minimum negative pressure for suction of the PMMA and the PDMS micro-suction tool, N_2 gas tank is used for the ignition source instead of the AIBN matrix [Fig. 7 (a)]. The fluid inlet of the punctured cover-plate is connected to an N_2 gas tank using a tube and a regulator. The negative pressure is measured with respect to the N_2 gas pressure from inlet using a pressure sensor (PSHK-760HAAG, SETech, Korea) connected to the suction area with a tube.

Fig. 7 (b) shows that the generated negative pressures of the PMMA and PDMS micro-suction tool are approximately linear to the increment of N_2 gas pressure. The PMMA micro-suction tool starts to pump the diluted ink when negative pressure is 0.11 kPa and inlet pressure is 5.0 kPa. The PDMS micro-suction tool starts to pump the diluted ink when negative pressure is 0.31 kPa and inlet pressure is 15.0 kPa. The minimum negative pressure of the PMMA micro-suction tool is approximately three times lower than that of the PDMS micro-suction tool because the contact angle



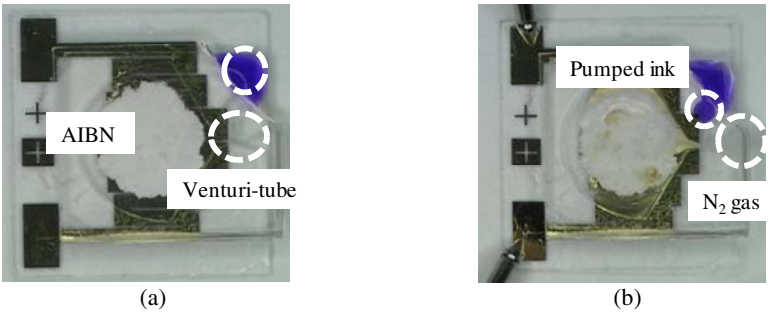
(a) (b)

Fig. 7. Negative pressure measurement of the PMMA and PDMS micro-suction tool (a) Measurement setup (b) Measurement result of the negative pressure

between PMMA and diluted ink is about 6 ° less than that of PDMS. This indicates that the PMMA micro-suction tool is operated with less amount of AIBN with efficiency.

3) Performance evaluation of the PMMA micro-suction tool

In Fig. 8 (a), the PMMA micro-suction tool using an AIBN propellant is implemented. The voltage is applied to the micro-heater using a power supply (E3620A, Agilent, America) for the AIBN ignition. The diluted ink in the suction area is successfully pumped into the reservoir, as shown in Fig. 8 (b). The pumped ink volume is measured to have 1.57 μL without any leakage of N_2 gas.



(a) (b)

Fig. 8. PMMA micro-suction tool using AIBN (a) Before the AIBN ignition (b) After the AIBN ignition

5 Conclusion

An advanced micro-suction tool fabricated with PMMA is presented. The bonding strength and minimum negative pressure for suction are measured to verify the performance of micro-suction tool. In bonding strength experiment, the PMMA Venturi-tube body is detached from glass substrate at 1.6 N/m. The measured value is five times larger than that of the PDMS micro-suction tool. The PMMA micro-suction tool starts to pump the diluted ink when negative pressure is 0.11 kPa. The minimum negative pressure for suction of the PMMA micro-suction tool is approximately three times lower than that of the PDMS micro-suction tool, which in turn results in an advanced suction efficiency. The fabricated PMMA micro-suction tool is able to pump 1.57 μL liquid into the reservoir without any leakage of N_2 gas. The capsular endoscope integrated with the PMMA micro-suction tool is expected to improve the diagnosis of the pathological research.

Acknowledgements

This research has been supported by the Intelligent Microsystem Center (IMC; http://www.microsystem.re.kr), which carries out one of the 21st century's Frontier R&D Projects sponsored by the Korea Ministry Of Commerce, Industry and Energy and Pohang Accelerator Laboratory (PAL; http://pls.postech.ac.kr).

References

1. Karagozler, M.E., Cheung, E., Kwon, J., Sitti, M.: Miniature Endoscopic Capsule Robot using Biomimetic Micro-Patterned Adhesives. In: Biomedical Robotics and Biomechatronics, Pisa, Italy, pp. 105–111 (2006)
2. Andrea, M., Arianna, M., Marc, O.S., Paolo, D.: Wireless capsule endoscopy: from diagnostic devices to multipurpose robotic systems. J. Biomed. Microdevices 9, 235–243 (2007)
3. Cavallotti, C., Piccigallo, M., Susiloa, E., Valdastri, P., Menciassia, A., Paolo, D.: An integrated vision system with autofocus for wireless capsular endoscopy. Sensors and Actuators A: Physical (2009)
4. Koo, K.I., Park, S.K., Ban, J.W., Cho, D.I.: A Novel Fluid Suction Tool For Capsular Endoscope. In: The 14th International Conference on Solid-State Sensors, Actuators and Microsystems, Lyon, France, vol. 1, pp. 1335–1336 (2007)
5. Hong, C.C., Murugesan, S., Kim, S.H., Beaucage, G., Choi, J.W., Ahn, C.H.: A Functional On-Chip Pressure Generator Using Solid Chemical Propellant for Disposable Lab-on-a-Chip. 3, 281–286 (2003)
6. Koo, K.I., Jeong, M.J., Park, S.K., Choi, H.M., Kim, K.S., Park, J.H., Cho, D.I.: Valveless, Venturi-tube Micro Suction Pump Using Solid Chemical Propellant. In: World Congress 2006, seoul, korea, pp. 306–309 (2006)

Statistical Analysis of Hippocampus Shape Using a Modified Mann-Whitney-Wilcoxon Test

Nikhil Ram Mohan[1], Carey Priebe[1], Youngser Park[1], and Majnu John[2]

[1] Center for Imaging Science, Johns Hopkins University,
3400 North Charles St., Baltimore, MD, 21218, USA
http://www.cis.jhu.edu
[2] Department of Biostatistics, Weill Cornell Medical College,
402 E. 67th St, New York, NY, 10065, USA
http://www.med.cornell.edu/publichealth/divisions/biostatistics/

Abstract. The Mann-Whitney-Wilcoxon (MWW) test statistic, while distribution-free, suffers from a loss of efficacy for certain underlying distributions. In this manuscript, we instead use a data-adaptive weighted generalized Mann-Whitney-Wilcoxon (AWGMWW) test statistic, one that is optimal in the Pitman Asymptotic Efficacy (PAE) sense, to discern differences in hippocampus shape among twin populations with or without Major Depressive Disorder (MDD). We verify, based on a previous study using the MWW statistic, that a high-risk group is more similar to the control group than the depressed group in terms of hippocampus shape. In addition, we show that the control group is more similar to the high-risk group than the depressed group - a distinction that could not be made in the preceding study. Our results suggest that the AWGMWW statistic is more powerful for this application.

Keywords: Hippocampus, Mann-Whitney-Wilcoxon, Major Depressive Disorder, Pitman Asymptotic Efficacy.

1 Introduction

1.1 Nonparametric Statistics

Wilcoxon and Mann-Whitney [1,2] are principally responsible for the advent of nonparametric statistics. The unbiased Rank-Sum and U-statistic have since been applied in a great deal of applications, and are most effective in situations where information about underlying distributions is not known *a priori*. One such application is in interpoint distance analysis [3] . The major obstacle in preventing even wider use of these statistics is the loss of efficacy owing to their nonparametric and distribution-free properties. To combat this, Xie & Priebe [4] provided weighted generalizations of the Mann-Whitney-Wilcoxon (WGMWW) and of the Wilcoxon-Signed-Rank (WGWSR) statistics, which were shown to be optimal in the Pitman Asymptotic Efficacy (PAE) sense. Unfortunately, these weighted generalizations were not practically viable owing to their containing parameters which were functions of the unknown null distribution. A data-adaptive

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 45–52, 2009.

alternative (AWGMWW) which has efficacy and power as good as WGMWW was provided by John & Priebe [5]. The efficacy and power of AWGMWW was compared with WGMWW and the classical MWW test statistic for several underlying densities. The results suggest that AWGMWW is only marginally more powerful and efficacious when the underlying density is normal or mildly left-skewed, but optimal when the underlying distribution is strongly right skewed, asymmetric bimodal or heavily kurtotic [5].

1.2 Major Depressive Disorder

Major Depressive Disorder (MDD), or Clinical Depression, is one of the most prevalent mental disorders in the United States, where approximately 16.2% of the population is affected at least once in their lifetime [6]. The disorder is characterized by a "combination of symptoms that interfere with an individual's ability to work, sleep, study, eat, and enjoy once-pleasurable activities" [7]. In order for an individual to be diagnosed with MDD, at least one of the primary symptoms, namely depressed mood or anhedonia, and at least three of the secondary symptoms, among marked weight loss, insomnia, fatigue, and thoughts of suicide, must be present for a period of six months or more [8]. The physiology of MDD has been examined in great detail over the past few decades. It is known that changes in the amounts of the neurotransmitters *serotonin* and *dopamine* are associated with depressive symptoms, and restoring the amounts of these neurotransmitters to normal levels is the function of most modern antidepressant drugs [9]. Neuroimaging studies have revealed that certain structural abnormalities such as "enlarged ventricles, sulci, or reduced volume of the frontal lobe and basal ganglia" are also significantly correlated with MDD [3,9].

1.3 Computational Anatomy

Computational Anatomy (CA) has only recently emerged as a discipline; it involves "the development of mathematical and software tools ... specialized to the study of brain anatomy" [10]. In fact, the shape of the brain as related to its anatomy is quite complex, and thus difficult to quantify. The underlying principle in CA is to "construct a mapping model"[3] that can measure the difference in shape between two brain regions-of-interest (ROIs). Several publications have used CA models on subsections of the brain to further the understanding of several psychological disorders; Posener et al. [11] examined the role of the hippocampus, a small structure found in the limbic system of the brain, in MDD, Miller et al. [12] studied the role of the cingulate gyrus in Dementia of the Alzhiemer's Type (DAT), while Csernansky et al.[13] correlated hippocampus shape changes with schizophrenia.

1.4 Goals

Recently, significant differences in hippocampus shape were gleaned between twin populations with or without Clinical Depression using the classical MWW test and interpoint comparison analysis [3]. The probability density function (pdf)

of the underlying distribution, simulated using a kernel estimator, was shown to be mildly right-skewed. The goal of this study is to develop an algorithm for the AWGMWW statistic and compare its efficacy and power with the classical MWW when applied towards interpoint comparison data from Park et al [3].

2 Data

Using Missouri birth records, three twin populations, including both monozygotic (MZ) and dizygotic (DZ) twin pairs, were recruited according to varying levels of Clinical Depression: (1) the Control group (CTRL) was unaffected by depression, (2) the High-Risk (HR) group included one twin in a pair having Depression and the other not, and the (3) Major Depressive Disorder (MDD) group having both twins in a pair being clinically depressed [3]. There were 59 twins (29 pairs, one unpaired) in CTRL, 22 twins in HR, and 33 twins (16 pairs, one unpaired) in MDD. All recruits were female, and a screening procedure excluded individuals with conditions that may influence structural changes of the brain, namely loss of consciousness greater than 5 minutes, pregnancy, and any chronic neurological illnesses [3].

High-resolution Magnetic Resonance Imaging (MRI) scans were obtained for all 114 individuals in the study, a first step in the analysis of hippocampus shape. Three MPRAGE scans ("160 slices at 256×256 FoV, 1 mm3 isotropic voxels") were acquired using the Siemens Vision/Sonata 1.5T scanner. Three-dimensional surfaces of the left and right hippocampi were extracted from the scans in accordance with current neuroanatomical guidelines [3].

3 Methods

3.1 Image Processing

The following protocol was the same as the one employed by Park et al [3]. For every hippocampus surface, 22 three-dimensional landmarks were anatomically defined. An "interpoint comparison matrix" $\tilde{D}$ was generated by applying a "non-parametric Landmark Matching (LM) transformation", which computes as a measure of shape difference the energy of the minimizing diffeomorphism between every possible pair of hippocampus landmarks [3].

For every pair of hippocampi, the following "error criterion" was used [3]:

$$\varphi^* = \arg\inf_{\varphi} \sigma d(\varphi_{\text{identity}}, \varphi)^2 + \sum_{i=1}^{N} ||\varphi(x_i) - y_i||^2$$

where "d is a geodesic distance in a group of diffeomorphisms and $\sigma > 0$ is a regularization parameter which controls the relative contribution of transformation complexity and the landmark mismatch to the optimization objective" [3]. The norm of the optimal mapping i.e., $LM(x, y) = ||\varphi^*||$, was used as the final measure of shape difference between the pair of hippocampi [3].

Such an analysis was conducted independently on the left hippocampi, the right hippocampi, and finally both hippocampi together. The resulting three interpoint comparison matrices were asymmetric, with dimension 114×114, and zeroes on the diagonal. Each of the matrices were symmetrized from $\tilde{D}$ to D using $d_{ij} = min\{\tilde{d}_{ij}, \tilde{d}_{ji}\}$, since the asymmetry was not an accurate reflection of hippocampus shape space. For more details, see Park et al [3].

3.2 Statistical Analysis

Stochastic Ordering. Each interpoint comparison matrix generated by the LM tranformation (see 3.1), being nonparametric, has an underlying probability density function (pdf) that is unknown. Park et al.[3] had shown, using "kernel probability density estimates" [14], that there exists a stochastic ordering relationship between the interpoint comparisons of HR and CTRL, written $d(HR, CTRL)$, and those of HR and MDD, written $d(HR, MDD)$, such that $d(HR, CTRL) <^{st} d(HR, MDD)$ [3]. Figure 1, a plot of the kernel pdfs, demonstrates this relationship.

Based on the structure of D, we conclude that every row of the interpoint comparisons corresponding to any one HR subject provides two vectors of comparison: one that compares that HR subject to every CTRL subject, and one that compares that HR subject to every MDD subject [3]. We call these two samples $\{d(HR_i, CTRL)\}$ and $\{d(HR_i, MDD)\}$ respectively; the HR subject's own twin is ignored when considering these samples in order to eliminate any bias that may arise from twinnedness not due to the underlying disease condition [3]. Park et al. [3] conducted a two-sample hypothesis test using the MWW statistic, where the null hypothesis was equality of the distributions of $d(HR, CTRL)$ and

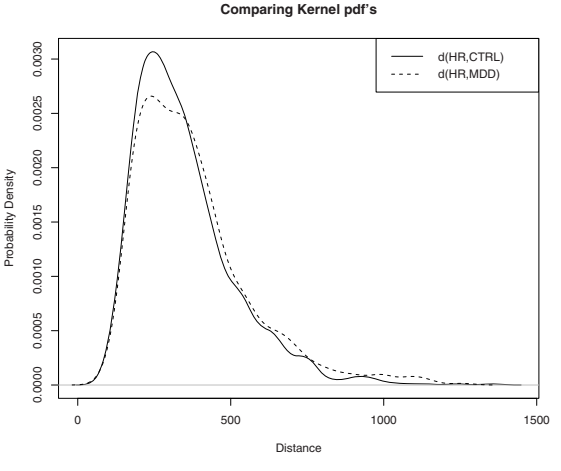


Fig. 1. This figure shows two kernel pdfs for $D_{LM-Left}$: the solid line shows $d(HR, CTRL)$ and the dashed line shows $d(HR, MDD)$ - this representation verifies the relationship obtained by Park et al.[3]

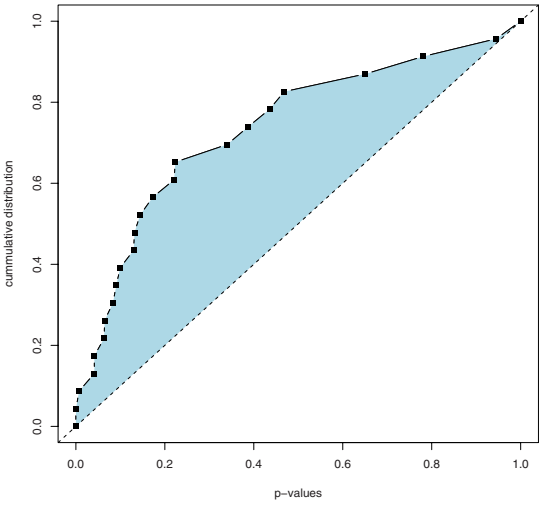


Fig. 2. QQ-plot for the two samples $d(HR, CTRL)$ and $d(HR, MDD)$ from $D_{LM-Left}$ using the MWW statistic - this verifies the distribution of p-values obtained by Park et al. [3]

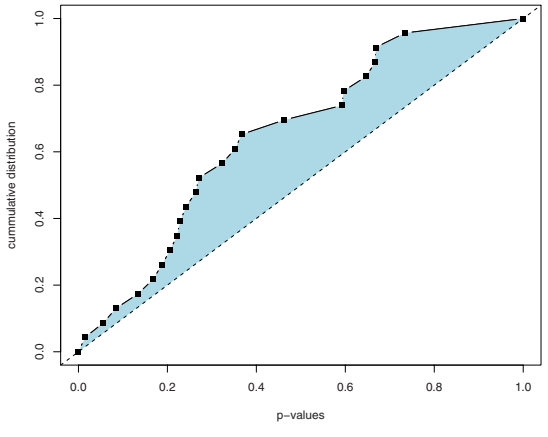


Fig. 3. QQ-plot for the two samples $d(HR, CTRL)$ and $d(HR, MDD)$ from $D_{LM-Left}$ using the AWGMWW statistic.

$d(HR, MDD)$ against the alternative of stochastic ordering [3]. Here, we instead use the AWGMWW statistic to conduct the same hypothesis test and obtain a p-value. Figure 2 and figure 3 compare the quantile-quantile plots of the p-values obtained by Park et al. using the MWW statistic [3], and those obtained in this study using the AWGMWW statistic, respectively.

If the null hypothesis were true, the p-values for both these plots should lie on the diagonal i.e. be distributed $Uniform(0, 1)$. Both plots, although with

different p-values, clearly suggest the alternative of stochastic ordering between the distributions. This seems to show that the modified statistic does not alter the overall result, but affects the strength of the underlying conclusion.

Classification. Recall, we have two samples $\{d(HR_i, CTRL)\}$ and $\{d(HR_i, MDD)\}$ for each HR subject, and that a p-value of 0.5 indicates that the distributions of these samples are equal. Thus, we can use the one-sided ($d(HR, CTRL) <^{st} d(HR, MDD)$) AWGMWW p-value (instead of the MWW p-value as implemented in Park et al [3]) to classify each HR subject as "closer" to CTRL or MDD. That is, if $p \leq 0.5$ for some HR subject, we classify that subject as CTRL, and if $p > 0.5$, we classify that subject as MDD. Extending this procedure to every HR subject would allow us to assess classifier performance on the HR group as a whole. In this case, we have fixed the HR group as the population of interest; we can equally fix one of the other groups (namely CTRL or MDD) as the population of interest, and follow the same procedure as above. Also note we can apply our classifier on the left hippocampi alone, the right hippocampi alone, or both left and right hippocampi simultaneously. The results are tabulated in Table 1.

Table 1. Classification of individual subjects based on AWGMWW p-values. The numbers in parentheses correspond to classification using the MWW statistic from Park et al [3]. The column-label "LM-Left" corresponds to the LM procedure being applied on the left hippocampi alone, and the row label "H:CvM" corresponds to the number of HR subjects classified as CTRL as opposed to MDD, and similarly, "H:MvC" corresponds to the number of HR subjects classified as MDD as opposed to CTRL. The rest of the column- and row-labels are defined analogously.

	LM-Left	LM-Right	LM-Left&Right
H:CvM	18(19)	16(16)	20(20)
H:MvC	4(3)	6(6)	2(2)
M:HvC	13(5)	14(9)	12(6)
M:CvH	20(28)	19(24)	21(27)
C:HvM	54(48)	27(22)	38(31)
C:MvH	5(32)	32(33)	21(25)

4 Results

Table 1 shows some interesting results. Firstly, we observe that the general classification trends for the AWGMWW statistic are consistent with the MWW statistic; for example, more HR subjects are classified as CTRL as opposed to MDD ($p < 0.00005$ for both MWW and AWGMWW respectively), and more MDD subjects are classified as CTRL than HR ($p < 0.00005$ for MWW and $p < 0.05$ for AWGMWW respectively).

However, we also notice some significant differences between the output of the MWW and the AWGMWW statistic. For example, using the MWW statistic on

$D_{LM-Left}$, more CTRL subjects were classified as HR against MDD ($p < 0.0005$), but on $D_{LM-Right}$, more CTRL subjects were classified as MDD against HR ($p < 0.0005$), and on $D_{LM-Left\&Right}$, the odds were approximately even between the two groups. This inconsistency was dubbed "a hemispheric ambiguity" by the authors of the original study [3].

Consider these revealing results from the present study; while the odds were again approximately even using the AWGMWW statistic on $D_{LM-Right}$, significantly more CTRL subjects were classified as HR as opposed to MDD when the statistic was used on $D_{LM-Left}$ and $D_{LM-Left\&Right}$ ($p \approx 0$ and $p < 0.01$ respectively). This appears to clear the "ambiguity" - CTRL is more like HR than MDD in terms of hippocampus shape.

5 Conclusions

The overarching results obtained using the AWGMWW statistic are consistent with those obtained using the MWW statistic, with an important addendum: HR is more like CTRL than MDD, MDD is more like CTRL than HR, and CTRL is more like HR than MDD. In the original study, Park et al.[3] concluded that CTRL was not like one group or the other in terms of hippocampus shape. Thus, we have demonstrated that AWGMWW, a PAE-optimal statistic, is indeed more powerful in this application.

Park et al.[3] had considered the relationship of the three twin populations as if hippocampus shape-space were one-dimensional. The authors had suggested that based on the classification results, the CTRL group was approximately in the middle between the HR and MDD groups in terms of hippocampus shape in $\mathbb{R}^1$ [3]. However, based on the novel result, it appears that although CTRL should be between HR and MDD, it must be closer to HR than it is to MDD. Thus, Figure 4 represents the revised "Artist's Rendition" that shows the relationship of the three populations in one dimension.

Fig. 4. Hippocampus shape-space if viewed in one dimension, a revision of the representation depicted by Park et al.[3].

References

1. Wilcoxon, F.: Individual Comparisons by Ranking Methods. Biometrics 1, 80–83 (1945)
2. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger that the other. Annals of Mathematical Statistics 18, 50–60 (1947)

3. Park, Y., Priebe, C.E., Miller, M.I., Mohan, N.R., Botteron, K.N.: Statistical Analysis of Twin Populations using Dissimilarity Measurements in Hippocampus Shape Space. Journal of Biomedicine and Biotechnology 2008, 5 pages, Article ID 694297 (2008), doi:10.1155/2008/694297
4. Xie, J., Priebe, C.E.: A weighted generalization of the Mann-Whitney-Wilcoxon statistic. Journal of Statistical Planning and Inference 102, 441–466 (2002)
5. John, M., Priebe, C.E.: A data-adaptive methodology for finding an optimal weighted generalized Mann-Whitney-Wilcoxon statistic. Computational Statistics & Data Analysis 51, 4337–4353 (2007)
6. Kessler, R.C., Bergland, P., Demler, O., Jin, R., Koretz, D., Merikangas, K., Rush, A.J., Walters, E.E., Wang, P.S.: The Epidemiology of Major Depressive Disorder: Results from the National Comorbidity Survey Replication (NCS-R). Journal of the American Medical Association 289, 3095–3105 (2003)
7. Depression: National Institute of Mental Health, http://www.nimh.nih.gov/health/publications/depression/complete-index.shtml
8. Diagnostic and Statistical Manual of Mental Disorders, 4th edn., American Psychiatric Association (1994)
9. Alloy, L.B., Riskind, J.H., Manos, M.J.: Abnormal Psychology: Current Perspectives, 9th edn., p. 278. McGraw-Hill Companies, New York (2005)
10. Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., Matejic, L.: Statistical Methods in Computational Anatomy. Statistical Methods in Medical Research 6, 267–299 (1997)
11. Posener, J., Wang, L., Price, J., Gado, M., Province, M., Miller, M., Babb, C., Csernansky, J.: High-dimensional mapping of the hippocampus in depression. American Journal of Psychiatry 160, 83–89 (2003)
12. Miller, M.I., Priebe, C., Qiu, A., Fischl, B., Kolasny, A., Brown, T., Park, Y., Busa, E., Jovicich, J., Yu, P., Dickerson, B., Buckner, R.L.: The Morphometry BIRN: Collaborative computational anatomy: The perfect storm for mri morphometry study of the human brain via diffeomorphic metric mapping. Human Brain Mapping 30, 2132–2141 (2008)
13. Csernansky, J., Joshi, S., Wang, L., Haller, J., Gado, M., Miller, J., Grenander, U., Miller, M.: Hippocampal Morphometry in schizophrenia by high dimensional brain mapping. Proceedings of the National Academy of Sciences 95, 11406–11411 (1998)
14. Silverman, B.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, New York (1996)

QTL Mapping for Forage Quality-Related Traits in Barley

Alireza Taleei[1,*], Barat Ali Siahsar[3], and Seyed Ali Peighambari[2]

[1,2] Professor and Associate Professor in the Department of Agronomy and Plant Breeding, Faculty of Science and Agricultural Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
[3] Assistant Professor in the Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Zabol, Iran
ataleei@ut.ac.ir

Abstract. Despite the importance of barley in animal feed, its forage quality has not been enough used as a selection criterion in breeding programs. To look for the genomic regions affecting barley forage quality, a population of 72 F_1-derived doubled haploid lines (DH) from the cross 'Steptoe/ Morex' and their two parents were planted in Karaj and Zabol provinces of Iran, in each under a randomized complete block arrangement with two replications. Forage samples were oven-dried and ground and dry matter digestibility (DMD), acid detergent fiber (ADF), neutral detergent fiber (NDF), acid detergent lignin, crude fiber (CF), crude protein (CP), water-soluble carbohydrates and ash content were measured by NIRS. Analyses of variance showed that genotype, environment and 'genotype – environment' interaction have significant effects on almost all studied traits. Several QTLs were resolved for each studied trait in both environments. Highest LOD scores were obtained for CF, ADF and DMD on chromosome 2H and for ash and CP on chromosomes 3H and 5H, respectively. QTLs for NDF were present on all chromosomes except 4H and 7H. 'QTL-environment' interaction and the specificity of these QTLs are discussed.

Keywords: QTL, 'Steptoe', 'Morex', forage quality, doubled haploids, stability, G x E interaction.

1 Introduction

Barley (Hordeum vulgare L.), the fourth most produced cereal worldwide, is mainly used for animal feed, human food and malting. It is also grown as a hay crop and has been shown to be an excellent forage source for livestock feed [1]. Traditionally, the main objectives of barley breeding programs are to develop cultivars with high-grain yield and high-malt quality. However, forage quality is an important selection criterion which is usually neglected in barley breeding programs. Based on nutritive parameters, forage should have optimum dry matter concentration for proper fermentation after ensiling, high digestibility and conversion efficiency to maximize intake, and high protein content to reduce requirements of supplemental protein [2]. Therefore, the quality of forage is influenced by several chemical factors including

* Corresponding author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 53 – 62, 2009.

dry matter digestibility (DMD), acid detergent fiber (ADF), neutral detergent fiber (NDF), acid detergent lignin (ADL), crude fiber (CF), crude protein (CP), water-soluble carbohydrate (WSC) and ash content [2-8].

Although forage quality can be evaluated directly by feeding experiments, but this method is costly and needs large quantities of breeding materials. Indirect assessment methods include; in vitro fiber digestibility, enzymatic digestion [9-11] and chemical analysis of cellular components [12]. However, these methods are slow and laborious, especially in case of multiple measurements for large samples. Near-Infrared Reflectance Spectroscopy (NIRS) offers an inexpensive, rapid, and accurate technique to evaluate forage-quality traits in breeding materials. The NIRS technique allows measurements of various forage quality traits to be made simultaneously, and is being used as a routine method for evaluating forage quality in many forage species [13, 14].

In barley, conventional breeding methods, which exploit substantial genetic variation and then select desirable traits, remain a successful approach for improving the forage quality traits. Direct phenotypic selection for high digestibility has been successful in several forage crop species [5], however, measuring methods for some traits like fiber, lignin and proteins are still time consuming and expensive. Forage quality traits, like most of other important traits, are quantitative in nature and controlled by several mild-effect genes, the so-called QTLs. With the rapid development of molecular marker technologies and preparation of dense linkage maps, it is now possible to identify markers harboring the QTLs of interest and use them in marker assisted selection programs. Information on the number and effects of QTLs can help breeders to better understand the genetic control of these traits and design more efficient selection strategies [15].

QTL analysis has been undertaken extensively in barley for different important traits. Genetic maps have been constructed from many different populations and employed to identify, locate and estimate the phenotypic effects of QTLs controlling economically important traits such as grain yield [16, 17], malting quality [18-21], disease resistance [22, 23] and environmental adaptation [24]. However, only a limited number of studies have been devoted to feed quality traits. Han et al. [7] performed a QTL analysis only for grain acid detergent fiber (ADF) and Abdel-Haleem et al. [25] mapped QTLs affecting feed quality-related traits such as grain ADF, grain starch, cracked grain particle size and cracked grain in situ dry matter digestibility.Breeding populations typically exhibit genotype by environment (G×E) interaction when tested in diverse environments. In such cases, at least some of the genes underlying QTLs would show ‘G×E’ interaction. This ‘QTL-E’ interaction would be expressed as significant effects detected only in a subset of environments or changes in the magnitude or even the sign of significant effects of the QTL across environments [26-29].

The main objectives of this study are: mapping of QTLs controlling forage quality traits in doubled haploid lines from the cross ‘Steptoe/ Morex’ and studies of their stability over different climatic conditions.

2 Materials and Method

Seventy-Two doubled haploid (DH) barley lines from the cross ‘Steptoe’ (CI15229) x ‘Morex’ (CI15773) together with both parents, were used to determine forage quality.

The DHs were developed through a modified 'Hordeum bulbosum' technique [30] by the Oregon State University Barley Breeding Program and were kindly provided by Dr. Patrick M. Hayes (Department of Crop and Soil Science, Oregon State University, Corvallis, OR 973314501, USA).

The 72 DHs as well as their parents were planted in two locations; the Research Farm of the Faculty of Agronomy and Animal Science of the University of Tehran in Karaj and the Research Farm of the Agricultural Research Station of Zabol in Zahak, under irrigated conditions. The experimental design was a randomized complete block with two replications. Each plot consisted of four rows; 3m long with row spacing of 25cm. Forage samples were taken at dough stage. A 0.5m clip sample from two middle rows was cut at stubble height and dried at 70˚c for 48 h. Dried samples were ground through the 0.1 mm screen of a cyclone mill and scanned using a Near Infrared Reflectance Spectroscopy (NIRS, Informatics 8600) with 6 - 20 wavelengths ranging from 500 to 2400nm.

Measurements were done on eight forage quality characteristics, namely dry matter digestibility (DMD), acid detergent fiber (ADF), neutral detergent fiber (NDF), acid detergent lignin (ADL), crude fiber (CF), crude protein (CP), water-soluble carbohydrates (WSC) and ash content. Analysis of variance was performed on the combined data from both locations using PROC GLM procedure in SAS [31]. Narrow-sense heritability was estimated as proposed by Singh et al. [32] for each trait. Moreover, simple correlations were calculated between the forage quality characteristics using entry means by the PROC CORR procedure in SAS [31].

For the current DH population, a molecular marker linkage map [33, 34] has been developed by the North American Barley Genome Mapping Project (currently at http://barleygenomics.wsu.edu/). This map comprises 327 markers with an average density of 3.75 cM [33, 34]. QTL analysis was conducted separately for each trait in each environment using Windows QTL Cartographer 2.5 [35]. For each trait, a series of 1000 permutations were run to determine the experiment-wise significance level, expressed as a LOD value, equivalent to P=0.05 [36]. Composite interval mapping (CIM) was employed to detect QTLs and estimate the magnitude of their effects [37, 38]. The genome was scanned at 2 cM intervals and the window size was set at 10 cM. Cofactors were chosen using the forward-backward stepwise regression. Confidence intervals of 95% were calculated by 1000-bootstrap re-sampling [39], as proposed in the Windows QTL Cartographer 2.5 package.

3 Results and Discussion

Analysis of variance of the 72 doubled haploid lines and their parents ('Steptoe' and 'Morex') showed that genotype had a highly significant ($P<0.01$) effect for all studied traits (results not shown). Also the simple effect of location and the genotype × location interaction were significant ($P<0.05$) for almost all studied traits. Previous studies have reported significant genotype and 'genotype - environment' interaction effects for different traits in this population [40, 41]. Also, similar variations within the 'Steptoe'/'Morex' DHs population for forage quality characteristics including grain starch, grain ADF, cracked particle size and in situ dry matter digestibility have been previously reported [7, 25].

'Morex' showed significantly higher values for DMD and ash, while ADF and CF had significantly higher values in 'Steptoe' whereas the difference between the two parents was not significant for other forage quality traits (results not shown). Han et al. [7] reported also that 'Steptoe' and 'Morex' had relatively high and low grain ADF content, respectively. The difference between parental mean and the mean of doubled haploids was not significant, except for DMD, ADF and CF, therefore, it is suggested that the 72 DHs of this study should be representatives of the possible DHs of this cross. Transgressive segregation, for both directions, were measured as a difference between superior parent and that of DHs with maximum trait value, and difference between inferior parent and the DHs with lowest trait value, were significant for DMD, ADF, NDF, ADL and WSC. For CF and ash content, positive and negative transgressive segregation were observed, respectively. In current population, transgressive segregation has been previously reported for different traits, including earliness, plant height, spike length, seeds per spike [17], grain protein, grain starch, grain ADF, grain particle size, dry matter, starch digestibility [25] and in-vitro regeneration ability [42]. Narrow-sense heritability estimates were high and ranged from 64.59 to 82.07% (results not shown). Therefore, all studied traits are highly heritable. Molina-Cano et al. [19] detected a high heritability for CF and Ray et al. [43] reported high heritability for in vitro dry matter digestibility, CP, NDF, ADF and hemicelluloses in crested wheatgrass, ranging from 50% up to 72%. These high heritability values suggest that selection for forage quality traits could be started in early generations. Apart from ash content and CP which didn't show any correlations with other traits, the rest of the studied traits were significantly correlated (results not shown). There was a negative relationship between DMD and ADF, NDF, ADL and CF. Digestible dry matter is often used as a proxy of digestible energy [2]. High fiber content (ADF, NDF and CF) and lignin (ADL) have negative impacts on animal intake, energy concentration, and digestibility of the forage [6, 44-47]. DMD had a positive relationship with WSC in our experiments. Accumulation of water-soluble carbohydrates is thought to be an important factor in forage quality, because it represents the most prominent energy source in finishing diets [48]. Multiple regression analysis of our data showed that 99.10% of DMD variation was expressed through ADF. The best fitted equation for DMD based on stepwise regression analysis was $DMD = 86.5760 + 7338ADF$. The same result has been reported in grasses, where DMD was accounted through ADF [45].

From one to five QTLs were detected for each of the eight forage quality traits, with phenotypic values (R2) between 6.5 to 26% (Table 4). Five QTLswere found only in Karaj and the other 6 QTLs were specific to Zabol. QTL models explained about 50.04 and 48.74% of total variation of DMD in Karaj and Zabol, respectively (Table 4). Three QTLs controlling variation in DMD were detected in both locations on chromosomes 2H, 3H and 5H. Zabol and Karaj each had a specific QTL for DMD on 1H. Abdel- Haleem et al., [25] reported that QTLs for DMD were located on chromosomes 1H, 2H, 3H and 6H. The coincidence of QTLs between Karaj and Zabol, and also between our- and the above-mentioned study, as well as the similarity of the signs and magnitudes of QTL effects between Karaj and Zabol, explains the high heritability estimate for DMD. There were four common QTLs identified on chromosomes 1H, 2H, 3H and 5H for ADF content. Abdel- Haleem et al. [25] reported four ADF QTLs on chromosomes 2H, 3H, 4H and 6H. Other studies

detected several QTLs controlling grain ADF content on 1H, 2H and 4H chromosomes in 'Steptoe/Morex' population [7, 49]. For NDF, two common QTLs on 1H and 3H, and a specific one on 2H in Karaj, as well as two specific QTLs in Zabol on 5H and 6H were found. The relatively high percentage of environment-specific QTLs implies the significance of 'genotype – environment' interaction for this trait, which is in accordance with the results reported by Fregeau-Reid et al. [50]. Three common QTLs were detected for ADL and just one specific in Zabol experiment. Chromosome 3H carries two QTLs for ADL. These two QTLs are far apart and also they have different additive signs. Positive allele for 'Qadl3Hasn' comes from 'Steptoe' and for 'Qadl3Hbsn' it comes from 'Morex'. Among the four QTLs detected for CF, the most prominent one is a common QTL with about 23% of phenotypic variance (R2) in both Karaj and Zabol experiments. Positive alleles for these QTLs, except the QTL on 3H, come from 'Morex' (Table 4). All of the QTLs for this trait were detected in both Karaj and Zabol with similar signs and R2s, which explains the non-significant 'genotype x environment' interaction. Molina-Cano et al. [19], also reported that this trait has a high heritability and doesn't show 'genotype – environment' interaction. Two common QTLs controlling variation in CP concentration were found on chromosomes 5H and 6H whereas two specific ones were found on 1H and 2H. Several grain protein QTLs have been detected on all of the seven barley chromosomes using different populations [15, 51, and 52].

Trait presents also the lowest heritability value among the studied traits (Table 2). A common QTL was mapped to 3H for WSC, and two environment-specific ones were mapped to 2H and 5H. A stable and quite strong QTL controlling ash was detected on 3H, while 1H and 4H each carried one specific QTL. Based on the results, chromosomes 2H and 3H present the greatest impact on forage quality.

Abdel- Haleem et al. [25], detected QTLs associated with grain protein content on 1H, 2H, 3H and 4H chromosomes. Being controlled by several non-stable QTLs, this in total, fifty-three QTLs were identified for the traits; forty-two of them were in common between Zabol and Karaj locations For example, the QTL for DMD ('Qdmd3Hsn') at the right hand of marker ABC171 on 3H coincided with QTLs for ADF, NDF, ADL, CF and WSC, (Table 1). QTLs 'Qadf2Hsn' for ADF and 'Qcf2Hsn' for CF were also co-localized. The same aspect is observed for location-specific QTLs. In Karaj, five QTLs for ash content, CP, ADF, DMD and CF were co-located on 1H at position about 60 cM. Peighambari et al. [17] also reported several co-locations for different traits, for example, QTLs for kernel weight and spikes per plant were located in the same region on 1H. These co-locations could be either because of linkage between two genes or the pleiotropy effect of one gene. In the case of pleiotropy, the correlation between traits will never be broken. Whatever, these coincidences provide genetic basis for the observed strong phenotypic correlations between the quality traits. The negative correlation between fiber components and DMD or WSC could have risen from co-locating QTLs with opposite effects. Thus, we may simultaneously decrease the fiber components and increase DMD or WSC through selection for either of the traits. It was proposed that incorporating only one fiber component in genetic mapping would be enough for altering fiber content of forage barley [6]. However, different traits affect various aspects of forage quality. For instance, NDF is negatively correlated to the intake potential of forage [44], but ADF is negatively related to the digestibility potential for the animal [45]. Therefore,

Table 1. Quantitative trait loci (QTLs) for forage quality traits in a 'Steptoe/Morex' doubled haploids population in Karaj and Zabol.

Trait	QTL name	Chromosome name	Nearest marker	QTL Position[a]	LOD score[b]		Additive effect		"R"	
					Karaj	Zabol	Karaj	Zabol	Karaj	Zabol
DMD	*Qdmd1Hsnk*	1H	Pcr2	63.1	2.51	-	0.63	-	6.92	-
	Qdmd1Hsnz	1H	His3B	100.0	-	2.63	-	0.66	-	7.87
	Qdmd2Hsn	2H	Adh8	62.0	6.22	5.00	1.04	1.13	20.37	21.05
	Qdmd3Hsn	3H	ABC171	24.3	3.56	3.96	-0.78	-0.91	11.09	13.49
	Qdmd5Hsn	5H	ABG705	36.7	6.66	4.27	1.00	0.94	18.58	14.20
ADF	*Qadf1Hsn*	1H	KsuF2A	59.0	3.78	3.00	-1.19	-1.05	12.01	9.47
	Qadf2Hsn	2H	Pox	52.6	4.65	5.96	-1.21	-1.48	14.39	19.61
	Qadf3Hsn	3H	ABC171	22.3	2.63	2.68	0.92	1.02	8.40	9.22
	Qadf5Hsn	5H	ABG705	36.7	4.70	4.45	-1.25	-1.29	15.75	14.81
NDF	*Qndf1Hsn*	1H	lca1	43.0	3.83	3.82	-1.45	-1.55	15.23	15.28
	Qndf2Hsnk	2H	ABG008	24.5	2.53	-	-0.84	-	7.36	-
	Qndf3Hsn	3H	ABC171	24.3	4.21	4.21	1.52	1.63	17.65	17.80
	Qndf5Hsnz	5H	ABG473	121.2	-	2.54	-	-0.13	-	6.44
	Qndf6Hsnz	6H	PSR167	1.5	-	2.94	-	-0.92	-	8.43
ADL	*Qadl1Hsn*	1H	ABG053	41.5	3.42	3.42	-0.11	-0.11	13.89	13.86
	Qadl3Hsn	3H	ABC171	24.3	3.29	3.28	0.10	0.11	12.82	12.78
	Qadl3Hbsn	3H	His4B	145.6	3.23	3.25	-0.15	-.15	14.13	14.24
	Qadl6Hsnz	6H	PSR167	1.4	-	2.54	-	-0.08	-	8.58
CF	*Qcf1Hsn*	1H	KsuF2A	59.0	3.27	3.36	-0.69	-0.75	9.68	9.94
	Qcf2Hsn	2H	Pox	52.6	6.80	6.88	-1.03	-1.11	22.67	22.90
	Qcf3Hsn	3H	ABC171	24.3	3.04	3.08	0.67	0.72	9.34	9.48
	Qcf5Hsn	5H	ABG705	36.7	3.49	3.36	-0.72	-0.75	10.70	10.20
CP	*Qcp1Hsnk*	1H	KsuF2A	59.0	2.99	-	-.40	-	10.64	-
	Qcp2Hsnz	2H	Pox	52.6	-	2.55	-	-0.57	-	9.57
	Qcp5Hsn	5H	WG364	103.8	4.82	5.20	0.32	0.36	17.95	19.77
	Qcp6Hsn	6H	ABR331	48.6	3.19	3.37	0.25	0.28	11.29	12.10
WSC	*Qwsc2Hsnz*	2H	ABG358	43.3	-	2.51	-	0.62	-	6.97
	Qwsc3Hsn	3H	ABC171	22.3	4.27	3.58	-0.68	-0.67	18.87	16.50
	Qwsc5Hsnk	5H	ABG705	36.7	2.59	-	0.53	-	7.36	-
Ash	*Qash1Hsnk*	1H	BCD351C	62.3	2.56	-	0.10	-	9.65	-
	Qash3Hsn	3H	MWG902	172.8	5.55	5.53	-0.33	-0.32	26.03	25.99
	Qash4Hsn	4H	cMWG652B	124.0	2.66	2.61	-0.17	-0.16	11.25	11.03

[a] QTL position expressed in cM, from origin of the linkage group (end of short arm).[b]Peak value of the LOD.[c] Percentage of phenotypic variance explained by the QTL**.** ***k***, Karj and *z*, Zabol locations, *dmd*, digestible dry matter; *adf,* acid detergent fiber*; ndf,* neutral detergent fiber; *adl* ,acid detergent lignin; *cf*, crude fiber; *cp*, crude protein; *wsc*, water-soluble carbohydrate.

several target traits would be necessary in marker assisted programs for improving barley forage quality.

One of the major goals of QTL mapping is to identify markers that are linked to genes affecting the trait of interest. QTL consistency across different environmental conditions is an important component for MAS, as would overcome problems associated with the interaction of QTL by environment. An important portion of forage quality QTLs which were detected in this study appears to be quite stable between locations; there is a promising scope for marker-assisted selection of these traits. In pervious studies, the utility of marker-assisted selection for traits such as yield and quality in barley has been demonstrated [41, 53]).

In this study we have focused on digestibility, protein, carbohydrate, mineral and fiber components traits, which are supposed to be associated with forage quality. A total of 21 common and 11 location-specific QTLs were found for the eight studied traits. The High number of common QTLs and the stability of their effects can increase the efficiency of marker assisted selection for forage quality traits in barley. Although the detected regions need to be mapped more precisely, tation obtained should help in marker-assisted selection.

References

1. Smith, W.C.: Barley. In: Crop Production, Evolution, History and Technology, pp. 174–291. John Wiley, New York (1995)
2. Coleman, S.E., Moore, J.E.: Feed quality and animal performance. Field Crops Res. 84, 17–29 (2003)
3. Lübberstedt, T., Melchinger, E.A., Klein, D., Degenhardt, H., Paul, C.: QTL mapping in test crosses of flint lines of maize: II. Comparison of different testers for forage quality traits. Crop Sci. 37, 1913–1922 (1997)
4. Smith, K.F., Reed, M., Foot, J.Z.: An assessment of the relative importance of specific traits for the genetic improvement of nutritive value in dairy pasture. Grass Forage Sci. 52, 167–175 (1997)
5. Casler, M.D.: Breeding forage crop for increased nutritive value. Adv. Agron. 71, 51–107 (2001)
6. Cardinal, A.J., Lee, M., Moore, K.J.: Genetic mapping and analysis of qualitative trait loci affecting fiber and lignin content in maize. Theor. Appl. Genet. 106, 866–874 (2003)
7. Han, F., Ullrich, S.E., Romagosa, I., Clancy, J.A., Froseth, J.A., Wesenberg, D.M.: Quantitative genetic analysis of acid detergent fiber content in barley grain. J. Cereal Sci. 38, 167–172 (2003)
8. Cogan, N.O.I., Smith, K.F., Yamada, T., Francki, M.G., Vecchies, A.C., Jones, E.S., Spangenberg, G.C., Forster, J.W.: QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass *(Lolium perenne L.)*. Theor. Appl. Genet. 110, 364–380 (2005)
9. DeBoever, J.L., Cottyn, F.X., Wainman, F.W., Vanacker, J.M.: The use of an enzymatic technique to predict digestibility, metabolisable and net energy of compound feedstuffs for ruminants. Anim. Feed Sci. Technol. 14, 203–214 (1986)
10. Mould, F.L.: Predicting feed quality-chemical analysis and in vitro evaluation. Field Crops Res. 84, 31–44 (2003)
11. Tessema, Z., Baars, R.M.T.: Chemical composition, in vitro dry matter digestibility and ruminal degradation of Napier grass *(Pennisetum purpureum (L.) Schumach.)* mixed with different levels of *Sesbania sesban (L.)* Merr. Anim. Feed Sci. Technol. 117, 29–41 (2004)
12. Jung, H.J.G.: Analysis of Forage Fiber and Cell Walls in Ruminant Nutrition. J. Nutr. 127, S810–S813 (1997)
13. Roberts, C.A., Workman, J., Reeves, J.B.: Near-infrared spectroscopy in agriculture. ASA-CSSA-SSSA, Inc, Madison WI (2004)
14. Mentink, R.L., Hoffman, P.C., Bauman, L.M.: Utility of near-infrared reflectance spectroscopy to predict nutrient composition and in vitro digestibility of total mixed rations. J. Dairy Sci. 89, 2320–2326 (2006)

15. Marquez-Cedillo, L.A., Hayes, P.M., Jones, B.L., Kleinhofs, A., Legge, W.G., Rossnagel, B.G., Sato, K., Ullrich, S.E., Wesenberg, D.M.: QTL analysis of malting quality in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theor. Appl. Genet. 101, 173–184 (2000)
16. Thomas, W.T.B., Baird1, E., Fuller, J.D., Lawrence, P., Young, G.R., Russell, J., Ramsay, L., Waugh, R., Powell, W.: Identification of a QTL decreasing yield in barley linked to Mlo powdery mildew resistance. Mol. breed. 4, 381–393 (1998)
17. Peighambari, S.A., Yazdi Samadi, B., Nabipour, A., Charmet, G., Sarrafi, A.: QTL analysis for agronomic traits in a barley doubled haploids population grown in Iran. Plant Sci. 169, 1008–1013 (2005)
18. Han, F., Ullrich, S.E., Kleinhofs, A., Jones, B.L., Hayes, P.M., Wesenberg, D.M.: Fine structure mapping of the barley chromosome 1 centromere region containing malt quality QTL. Theor. Appl. Genet. 95, 903–910 (1997)
19. Molina-Cano, J.L., Francwsch, M., Perez-Vendrell, A.M., Ramo, T., Voltas, J., Brufau, J.: Genetic and environmental variation in malting and feed quality of barley. J. Cereal Sci. 25, 37–47 (1997)
20. Borem, A., Mather, D.E., Rosmusson, D.C., Fulcher, R.G., Hayes, P.M.: Mapping quantitative trait loci for starch granule traits in barley. J. Cereal Sci. 29, 153–160 (1999)
21. Beecher, B., Smidansky, E.D., See, D., Blake, T.K., Giroux, M.J.: Mapping and sequence analysis of barley hordoindolines. Theor. Appl. Genet. 102, 833–840 (2001)
22. Chen, F., Prehn, D., Hayes, P.M., Mulrooney, D., Corey, A., Vivar, H.: Mapping genes for resistance to barley stripe rust *(Puccinia striiformis f. sp. horde.)*. Theor. Appl. Genet. 88, 215–219 (1994)
23. Steffenson, B.J., Hayes, P.M., Kleinhofs, A.: Genetics of seedling and adult plant resistance to net blotch *(Pyrenophora teres f. sp. teres)* and spot blotch (Cochliobolus sativus) in barley. Theor. Appl. Genet. 92, 552–558 (1996)
24. Mickelson, S., See, D., Meyer, F.D., Garner, J.P., Foster, C.R., Blake, T.K., Fischer, A.M.: Mapping of QTL associated with nitrogen storage and remobilization in barley *(Hordeum vulgare L.)* leaves. J. Exp. Bot. 54, 801–812 (2003)
25. Abdel-Haleem, H., Giroux, M., Talbert, H., Bowman, J., Kanazin, V., Blake, T.: Identification of QTLs controlling the feed quality of barley. In: Plant & Animal Genome XII Conf., San Diego, CA, January 10-14, p. 468 (2004)
26. Marquez-Cedillo, L.A., Hayes, P.M., Kleinhofs, A., Legge, W.G., Rossnagel, B.G., Sato, K., Ullrich, S.E., Wesenberg, D.M.: QTL analysis of agronomic traits in barley based on the doubled haploid progeny of two elite North American varieties representing different germplasm groups. Theor. Appl. Genet. 103, 625–637 (2001)
27. Ungerer, M.C., Halldorsdottir, S.S., Purugganan, M.D., Mackay, T.F.: Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. Genetics 165, 353–365 (2003)
28. Malosetti, M., Voltas, J., Romagosa, I., Ullrich, S.E., van Eeuwijk, F.A.: Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137, 139–145 (2004)
29. Mayo, O.: Interaction and quantitative trait loci. Aust. J. Exp. Agric. 44, 1135–1140 (2004)
30. Chen, F., Hayes, P.M.: A comparison of *Hordeum bulbosum*-mediated haploid production efficiency in barley using in vitro floret and tiller culture. Theor. Appl. Genet. 77, 701–704 (1989)
31. SAS Institute, SAS State user's guide 9.1: Statistics. SAS Inst., Cary, NC (1992)
32. Singh, M., Ceccarelli, S., Hamblin, J.: Estimation of heritability from varietal trials data. Theor. Appl. Genet. 86, 437–441 (1993)

33. Kleinhofs, A., Kilian, A., Saghai Maroof, R.M., Biyashev, R.M., Hayes, P.M., Qchen, F., Laption, N., Fenwick, A., Blake, T.K., Kanazin, V., Ananiev, E., Dahleen, L., Kudrna, D., Bollinger, J., Knapp, S.J., Liu, B., Sorrells, M., Heun, M., Franckowiak, J.D., Hoffman, D., Skaden, R., Steffeson, B.J.: A molecular, isozyme and morphological map of the barley *(Hordeum vulgare L.)* genome. Theor. Appl. Genet. 86, 705–712 (1993)
34. Hayes, P.M., Liu, B.H., Knapp, S.J., Chen, F., Jones, B., Blake, T., Franckowiak, J., Rasmmusson, D., Sorrells, M., Ullrich, S.E., Wesenberg, D., Kleinhofs, A.: Quantitative trait locus effects and environmental interaction in a Sample of North American barley germplasm. Theor. Appl. Genet. 87, 392–401 (1993)
35. Wang, S., Basten, C.J., Zeng, Z.B.: Windows QTL cartographer 2.5, Department of Statistics, North Carolina State University, Raleigh, NC (2007), `http://statgen.ncsu.edu/qtlcart/wQTLcart.htm`
36. Churchill, G.A., Doerge, R.W.: Empirical threshold values for quantitative trait mapping. Genetics 138, 963–971 (1994)
37. Jansen, R.C., Stam, P.: High resolution of quantitative traits into multiple loci via interval mapping. Genetics 138, 1447–1455 (1994)
38. Zeng, Z.B.: Precision mapping of quantitative trait loci. Genetics 136, 1457–1468 (1994)
39. Lebreton, C.M., Visscher, P.M., Haley, C.S., Semikhodskii, A., Quarrie, S.A.: A nonparametric bootstrap method for testing close linkage vs. pleiotropy on coincident quantitative trait loci. Genetics 150, 931–943 (1998)
40. Romagosa, I., Han, F., Ullrich, S.E., Hayes, P.M., Wesenberg, D.M.: Verification of yield QTL through realized molecular marker- assisted selection responses in a barley cross. Mol. Breed. 5, 143–152 (1999)
41. Zhu, H., Briceno, G., Dovel, R., Hayes, P.M., Liu, B.H., Liu, C.T., Ullrich, S.E.: Molecular breeding for grain yield in barley: an evaluation of QTL effects in a spring barley cross. Theor. Appl. Genet. 98, 772–779 (1999)
42. Bregitzer, P., Campbell, R.D.: Genetic markers associated with green and albino plant regeneration from embryogenic barley callus. Crop Sci. 41, 173–179 (2001)
43. Ray, I.M., Karn, J.F., Dara, S.T.: Heritabilities of nutritive quality factors and interrelationships with yield in tetraploid crested wheatgrass. Crop Sci. 36, 1488–1491 (1996)
44. Mertens, D.R.: Predicting intake and digestibility using mathematical models of ruminal function. J. Anim. Sci. 64, 1548–1558 (1987)
45. Reid, R.L., Jung, G.A., Thayne, W.V.: Relationships between nutritive quality and fiber components of cool season and warm season forages: A retrospective study. J. Anim. Sci. 66, 1275–1291 (1988)
46. McDonald, P., Edwards, R.A., Greenhalgh, J.F.D., Morgan, C.A.: Animal Nutrition, 5th edn. Longman Scientific & Technical, New York (1995)
47. Perry, T.W., Cullison, A.E., Lowrey, R.S.: Feeds and Feeding, 5th edn. Prentice Hall, New Jersey (1999)
48. Newman, C.W., Newman, R.K.: Nutritional aspect of barley seed structure and composition. In: Shewry, P.R. (ed.) Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, CAB International, pp. 351–368 (1992)
49. Ullrich, S.E., Han, F., Froseth, J.A., Jones, B.L., Newman, C.W., Wesenberg, D.M.: Mapping of loci that affect carbohydrate content in barley grain. In: Slinkard, A., Scoles, G., Rossnagel, B. (eds.) Proceedings of the V Int'l Oat Conf. & VII Int'l Barley Genet. Symp. Poster Ses., vol. 1, pp. 141–143. Univ. of Saskatchewan Ext. Press, Saskatoon (1996)

50. Frègeau-Ried, J., Choo, T.M., Ho, K.M., Martin, R.C., Konishi, T.: Comparisons of two-row and six-row barley for chemical composition using doubled-haploid lines. Crop Sci. 41, 1737–1743 (2001)
51. Oziel, A., Hayes, P.M., Chen, F.Q., Jones, B.: Application of quantitative trait locus mapping to the development of winter-habit malting barley. Plant Breed 115, 43–51 (1996)
52. Mather, D.E., Tinker, N.A., LaBerge, D.E., Edney, M., Jones, B.L., Rossnagle, B.G., Legge, W.G., Briggs, K.G., Irvine, R.B., Falk, D.E., Kasha, K.J.: Regions of the genome that affect grain and malt quality in a North American two-row barley cross. Crop Sci. 37, 544–554 (1997)
53. Ayoub, M., Armstrong, E., Bridger, G., Fortin, M.G., Mather, D.E.: Marker-based selection in barley for a QTL region affecting alpha amylase activity of malt. Crop Sci. 43, 556–561 (2003)

Fluorimetric Determination of L-3-Hydroxybutyrate Concentrations in the Serum of Normal and Aristolochic Acid-Treated Mice

Chien-Ming Chen[1], Yih-Huei Uen[2, *], Chen-Yi Kuo[2], Tzu-Chuan Huang[3], and Jen-Ai Lee[3, **]

[1] Department of Electro-optical Engineering, National Taipei University of Technology, No.1, Sec. 3, Chung-Hsiao E. Rd. Taipei 106, Taiwan
[2] Department of Surgery, Chi-Mei Hospital, Tainan, Taiwan
No. 901, Chung-Hwa Rd. Yong Kang City, Tainan 710, Taiwan
[3] Department of Pharmaceutical Analysis, School of Pharmacy, Taipei Medical University, No. 250, Wu-Hsing St., Taipei 110, Taiwan
Fax: +886-2-2736-1661; Ext. 6125
jenai@tmu.edu.tw

Abstract. We determined the concentrations of L-3-hydroxybutyrate (L-3HB) in the serum of normal and aristolochic acid (AA)-treated mice. The samples were derivatized with 4-nitro-7-piperazino-2,1,3-benzoxadiazole (NBD-PZ) and determined fluorimetrically by column-switching high-performance liquid chromatography (HPLC). Total 3-HB derivatives in serum were separated by an ODS column and then introduced to a sample loop for enantiomeric separation by two tandem CHIRALCEL OD-RH columns. The results showed that the concentration of L-3HB in the serum of normal mice was 0.81 ± 0.15 μM compared to that of AA-treated mice with 4.33 ± 2.81 μM after 3 days and 8.65 ± 2.60 μM after 5 days of AA administration ($p < 0.05$), suggesting that serum L-3HB may be used as an indicator to determine the level of renal damage.

Keywords: L-3-hydroxybutyrate; fluorimetric determination; renal damage; column-switching HPLC.

1 Introduction

The 3-HB has a chiral center at the third carbon, and thus exists as two enantiomers. It has long been believed that the enantiomer of D-3HB, L-3HB, is absent in mammalian tissues or body fluids under physiological conditions. However, recent reports have demonstrated free L-3HB in a variety of rat tissues, such as serum, liver, heart, and kidney [1, 2]. It was reported that L-3HB can also be used the same as D-3HB in the biosynthesis of hepatic lipids, brain proteins, and amino acids in neonatal rats [3, 4]. Furthermore, L-3HB was shown to be a more-favorable substance than D-3HB,

* Co-first author.
** Corresponding author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 63–68, 2009.

acetoacetate or acetone for sterol and fatty acid synthesis in the brain, spinal cord, and kidney [5]. Rho et al. reported that neither D-3HB nor (D+L)-3HB, but only L-3HB showed anticonvulsant effect in the animal studies [6]. The effect of L-3HB may be via the block of NMDA receptors [7]. In addition, in patients with β-ketothiolase deficiency or medium-chain acyl-CoA dehydrogenase deficiency, L-3HB was found to be comprised of a minor amount of about 3–5% of the total (D+L)-3HB by gas chromatography-mass spectrometry [8]. The results suggest that the differences in the enantiomeric ratio of 3HB may originate from the enantioselectivity of different enzyme systems [8]. Therefore, the purpose of this study was to determine L-3HB concentrations in the serum of normal and AA-induced mice in order to elucidate whether L-3HB is an adequate indicator of renal damage.

2 Materials and Methods

2.1 Chemicals

NBD-PZ, triphenylphosphine (TPP), and 2,2'-dipyridyl disulfide (DPDS) were purchased from Tokyo Kasei Kogyo (Tokyo, Japan). Sodium D- and L-3HB were obtained from Wako Pure Chemicals (Osaka, Japan). Trifluoroacetic acid (TFA) was from Riedel-de Haën (Seelze, Germany), and propionic acid was from Nacalai Tesque (Kyoto, Japan). HPLC-grade acetonitrile (MeCN), ethanol (EtOH), and methanol (MeOH) were the products of the Merck Co. (Darmstadt, Germany). TSKgel ODS-80Ts (150 × 4.6 mm i.d.) and CHIRALCEL OD-RH (OD-RH) (150 × 4.6 mm i.d.) were from the Tosoh Co. (Tokyo, Japan) and Daicel Co. (Osaka, Japan), respectively.

2.2 Derivatization of D- and L-3HB with NBD-PZ

The dericatization of D- and L-3HB with NBD-PZ were carried out according to our previous method [1]. Briefly, One hundred microliters of the sample solution was added to 100 μL of 2 mM NBD-PZ dissolved in MeCN, and then 50 μL each of 280 mM TPP and DPDS in MeCN was added and mixed. The derivatization was allowed to stand for 3 h at 30°C; followed by addition of 250 μL of 0.1% TFA in H_2O to terminate the reaction. An Empore^{TM} SDB-RPS cartridge (3M, St. Paul, MN, USA) was preconditioned with 100 μL of an eluting solution composed of EtOH/MeCN/H_2O (20/30/50, v/v/v), and the eluent was discarded. We than loaded 100 μL of the resultant solution onto the cartridge and collected the eluent. Another elution was performed by loading 100 μL of the eluting solution to thoroughly elute D- and L-3HB. Two portions of the eluent were combined and filtrated through 0.20-μm filters (Sartorius AG, Göttingen, Germany), and 20 μL of the filtrate was injected into the HPLC.

2.3 Preparation of Mouse Serum

Animal experiments were approved by the Laboratory Animal Research Committee of Taipei Medical University. Serum used for L-3HB analysis was from 6-week-old C3H/He mice (National Laboratory Animal Breeding and Research Center, Taipei,

Taiwan), which were reared in cages with food and tap water provided ad libitum. To induce renal damage, aristolochic acid sodium salt dissolved in distilled water was administered orally at a dose of 3.0 μg/mL (0.5 mg/Kg/day). After 3 and 5 days, the blood of experimental animals from normal and AA-treated mice was drawn from the tail vein and centrifuged with Mikro 22R Centrifuge (Hettich Zentrifugen, Tuttlingen, Germany) at 700 g for 15 min at 4°C immediately. Ten microliters of 1.0 mM propionic acid in H_2O as an internal standard (I.S.) was added to 50 μL of serum, and was then brought to 200 μL with EtOH for deproteinization. The solution was vigorously mixed using a vortex mixer and centrifuged at 700 g for 5 min. One hundred microliters of the supernatant was supplemented with fluorogenic reagents and TPP and DPDS to perform the derivatization as described above.

2.4 HPLC Conditions

A column-switching HPLC system, as described in our previous publications [1, 2, 9], was used in the present work. As illustrated in Fig. 1, the HPLC system was equipped with a AS-950 intelligent sampler (Jasco, Tokyo, Japan) with a 20-μL loop, two pumps (L-7100; Hitachi, Tokyo, Japan), an L-2480 and an L-2485 fluorescence detector (Hitachi), two D-2500 chromatointegrators (Hitachi), and a Rheodyne Model 7000 switching valve with a 100-μL loop as a sample trap. A TSKgel ODS-80Ts column was used for separation and quantification of the total 3HB in mouse serum; the mobile phase was MeOH/H_2O (33/67, v/v) at a flow rate of 0.7 mL/min. Two OD-RHs connected in tandem (tandem OD-RHs) were used for the enantiomeric separation of D- and L-3HB eluted with MeCN/H_2O (40/60, v/v); the flow rate was 0.3 mL/min. Fluorescence detection (detectors 1 and 2) was performed at 547 nm with a 491 nm excitation wavelength.

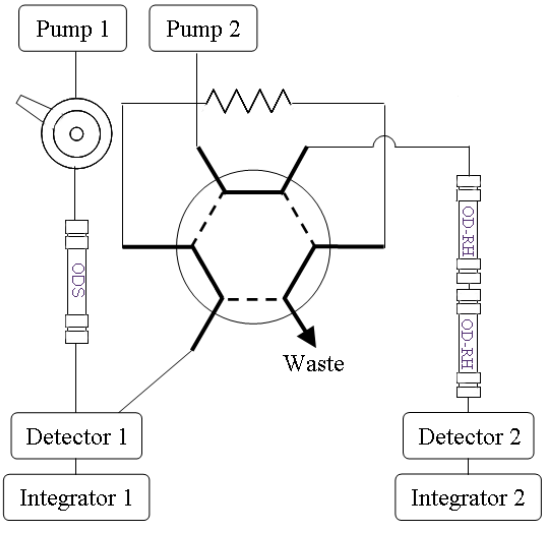


Fig. 1. A schematic diagram for the column-switching HPLC system used in this study. The solid and dotted lines indicate six-port valve position A and B, respectively.

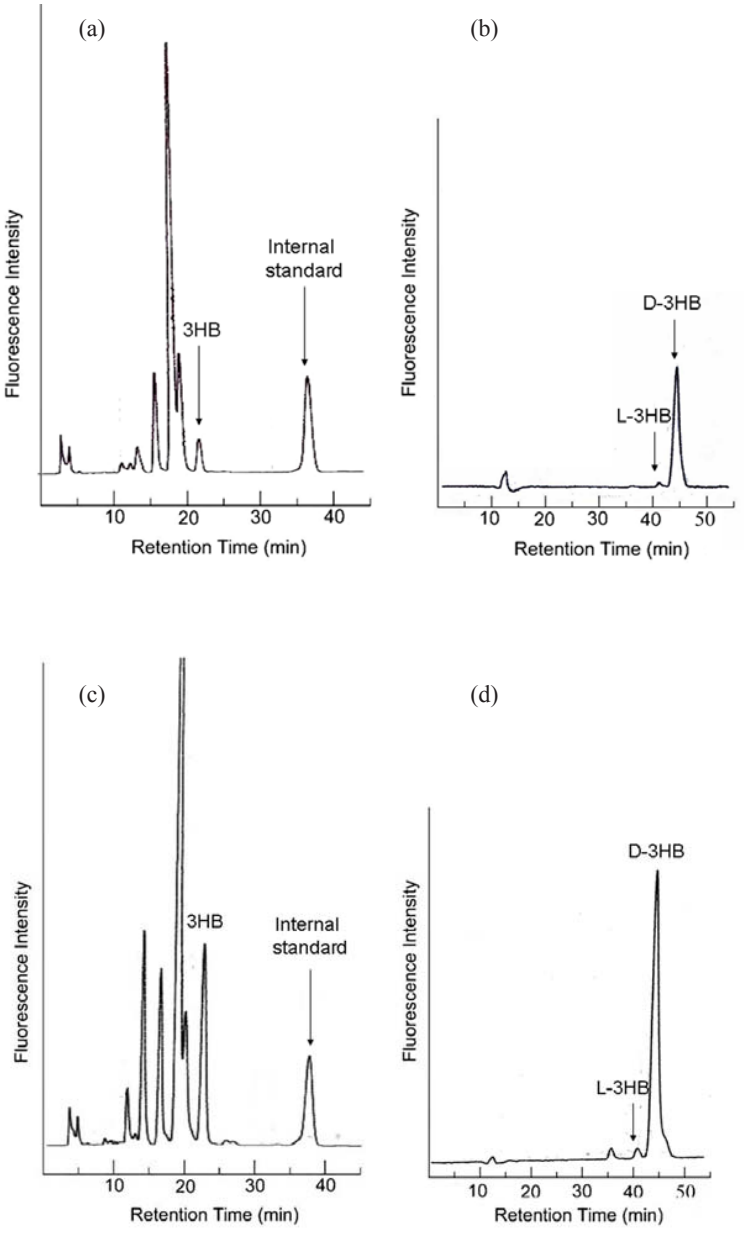


Fig. 2a-d. Representative chromatograms obtained by proposed column-switching HPLC system: (a) Analysis of L-3HB in normal mouse serum. The retention times of the total 3HB derivative and the I.S. were about 23 and 38 min, respectively. (b) Enantiomeric separation of D- and L-3HB derivatives isolated from mouse serum. (c) (A) Analysis of L-3HB in AA-treated mouse serum. (B) Enantiomeric separation of D- and L-3HB derivatives isolated from AA-treated mouse serum.

3 Results and Discussion

By using the proposed column-switching HPLC method, D- and L-3HB concentrations in the serum of normal and AA-treated mice were investigated. The derivatized total 3HB could be separated from other carboxylate components in mouse serum using a TSKgel ODS column with isocratic elution of MeOH/H_2O (33/67) at a flow rate of 0.7 mL/min. The peak of total 3HB on the chromatogram emerged at about 23 min, and the retention time of the internal standard was about 38 min (Fig. 2a, 2c). Neither the total 3HB nor the internal standard peaks were found to overlap with any other interfering peaks, and quantification of the total 3HB concentration was accomplished through the calibration curve.

For the determination of L-3HB in mouse serum, an octylsilica column (TSKgel ODS-80Ts) was selected as the first non-chiral column, because it gave good separation of 3HB from the other endogenous compounds in mouse serum (Fig. 2a and c). After the separation on the octylsilica column, a portion of 3HB fraction was introduced into the cellulose-based chiral column through a six-port valve. As a result, both peaks of D-3HB and L-3HB were clearly observed in the chromatogram (Fig. 2b and d). A significant increase of total 3HB was observed in the serum of AA-treated mice. The results are shown in Fig. 3. In normal mice, the concentrations of L-3HB in serum was 0.81 ± 0.15 μM (mean ± SD; 3days) and 0.86 ± 0.81 μM (5 days), while those in serum of AA-treated mice were 4.33 ± 2.81 μM (3 days; $p<0.01$) and 8.65 ± 2.60 μM (5 days; $p<0.05$), respectively. Not only L-3HB, but also D-3HB in serum, was significantly increased in the AA-treated mice (data not shown). The average L-3HB percentage was about 2.88% of the total 3HB in normal mouse serum, and that was about 2.03% in AA-treated ones. To our knowledge, the present work is the first study of the identification and quantification of L-3HB under disease conditions.

In conclusion, employing the column-switching HPLC system, we were able to detect and identify L-3HB in normal and AA-treated mouse serum. A significant increase of L-3HB was observed in the serum of AA-treated mice. It suggests that serum L-3HB may be used as an indicator to determine the level of renal damage.

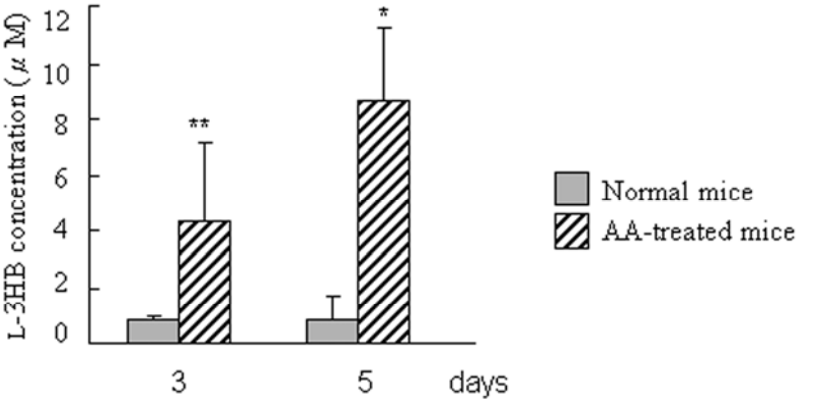


Fig. 3. Concentrations of L-3HB in the serum of normal and AA-treated mice after 3 and 5 days of AA administration. ($n = 5$). *$p < 0.05$, **$p < 0.01$ vs normal rats.

Acknowledgments. This work was financially supported by the Chi Mei Medical Center (95CM-TMU-16).

References

1. Tsai, Y.C., Liao, T.H., Lee, J.A.: Identification of L-3-hydroxybutyrate as an original ketone body in rat serum by column-switching high-performance liquid chromatography and fluorescence derivatization. Anal. Biochem. 319, 34–41 (2003)
2. Tsai, Y.C., Chou, Y.C., Wu, A.B., et al.: Stereoselective effects of 3-hydroxybutyrate on glucose utilization of rat cardiomyocytes. Life Sci. 78, 1385–1391 (2006)
3. Swiatek, K.R., Dombrowski, G.J., Chao, K.L., Chao, H.L.: Metabolism of L- and D-3-hydroxybutyrate by rat liver during development. Biochem. Med. 25, 160–167 (1981)
4. Swiatek, K.R., Dombrowski, G.J., Chao, K.L.: The metabolism of D- and L-3-hydroxybutyrate in developing rat brain. Biochem. Med. 31, 332–346 (1984)
5. Webber, R.J., Edmond, J.: Utilization of L(+)-3-hydroxybutyrate, D(-)-3-hydroxybutyrate, acetoacetate, and glucose for respiration and lipid synthesis in the 18-day-old rat. J. Biol. Chem. 252, 5222–5226 (1977)
6. Rho, J.M., Anderson, G.D., Donevan, S.D., White, H.S.: Acetoacetate, acetone, and dibenzylamine (a contaminant in L-(+)-b-hydroxybutyrate) exhibit direct anticonvulsant actions in vivo. Epilepsia 43, 358–361 (2002)
7. Donevan, S.D., White, H.S., Anderson, G.D., Rho, J.M.: Voltage-dependent block of N-methyl- D-aspartate receptors by the novel anticonvulsant dibenzylamine, a bioactive constituent of L-(+)-beta-hydroxybutyrate. Epilepsia 44, 1274–1279 (2003)
8. Heil, M., Podebrad, F., Prado, E., et al.: Enantioselective analysis of ketone bodies in patients with b-ketothiolase deficiency, medium-chain acyl coenzyme A dehydrogenase deficiency and ketonemic vomiting. J. Chromatogr. B 739, 313–324 (2000)
9. Lee, J.A., Tsai, Y.C., Chen, H.Y., et al.: Fluorimetric determination of D-lactate in urine of normal and diabetic rats by column-switching high-performance liquid chromatography. Anal. Chim. Acta 534, 185–191 (2005)

Designing a Column-Switching High-Performance Liquid Chromatograph System for Enantiomeric Separation of Mouse Urinary D,L-Lactate

Chien-Ming Chen* and Chi-Fu Yen

Department of Electro-Optical Engineering, National Taipei University of Technology,
No.1, Sec. 3, Chung-Hsiao E. Rd.
106 Taipei, Taiwan
cmchen@ntut.edu.tw

Abstract. We designed a circuit for turning the switching valve automatically, which is including mechanical units of motion, two timers, and electronic control units, though it is to combine two high-performance liquid chromatograph (HPLC) systems with a Rheodyne Model 7000 switching valve for separation of D- and L-lactate in biological samples has been reported in previous studies. The HPLC system comprised two HPLCs with an autosampler and two integrators. For more effective analysis, Total (D+L)-lactate fluorescent derivatives was isolated by a TSKgel ODS-80Ts column, when the peak of lactate derivatives appeared in the chromatogram on the ODS column, the valve position was changed by switching and then the isolated lactate derivatives was introduced into the chiral column automatically. Then the D- and L-enantiomers were separated by the chiral column. The utility of the analyzer system was tested for the determination of enantiomeric separation of urinary D, L-lactate in the mouse. It takes about 90 min to analyze one urine sample for a circulation. Before automation, we must turn the switching valve manually when the signal shows up, so that there must be someone always standby and only about 10 samples could be analyzed within one day. With the designed automatic column-switching HPLC system, the switching valve can be automatically turned at the setting time repeatedly and over 15 samples could be determined within one day. The automatic system makes the analysis of D-lactate be time-saving, convenient and reproducible.

Keywords: column-switching HPLC; switching valve; automation; enantiomeric separation; D-lactate.

1 Introduction

High-performance liquid chromatography (HPLC) is a popular analytical method and is commonly used for the separation of chemical compounds. For investigating the role of D-lactate in mammals, we developed highly sensitive HPLC methods for determination of D-lactate in biological samples like rat serum [1-3], urine [4], and human serum [5]. The proposed methods were further applied to the plasma [2] or

* Corresponding author.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 69–75, 2009.

urine [4] of diabetic rats induced by intraperitoneal administration of streptozotocin, and the significant increases of D-lactate concentrations was observed in the diabetic rats as compared to the normal ones. In diabetic rats, D-lactate concentrations revealed a rising trend from the 7th day and then kept stable from the 28th day after induction, suggesting that urinary D-lactate may be used as a marker to determine the diabetic stage and the level of kidney damage [4]. Significantly increased D-lactate and L-lactate concentrations were also observed in the serum of diabetic patients as compared with normal subjects [5]. As mentioned above, the HPLC methods are effective in analyzing D-lactate concentration under physiological or pathological conditions.

All the HPLC systems used in the above methods contained two HPLCs with a switching valve, as illustrated in Fig. 1. Total (D+L)-lactate fluorescent derivatives was isolated by an ODS column in first HPLC and then the isolated lactate derivatives was introduced into the chiral column by rotating the switching valve manually. Then the D- and L-enantiomers were separated on the chiral column. It takes 34 min to obtain the total (D+L)-lactate, and an extra 20 min for washing column and another 30 min for returning to the initiate sate. It takes average time of one and a half hours to complete a separation circle. If the designed automatic valve were not set in the separation system, the switching valve must be turned manually when the signal shows up, so that there must be someone always standby and only about 10 samples could be analyzed within one day at most. Although there are many two-dimensional HPLC systems in the market, all of them are the type of time setting. In the study, we desired to construct a simple, cheap automatic switching valve that can be used in our existing HPLC system.

In the present study, we designed an automation system to rotate the switching valve when total (D+L)-lactate is just isolated, and drive the second integrator to record the chromatogram. In Fig. 1 the relevant coordinate system and parameters of a valve are shown. With the designed automatic column-switching HPLC system, the investigation of the correlation between D-lactate concentration in mouse urine and the stage of kidney damage will speed up.

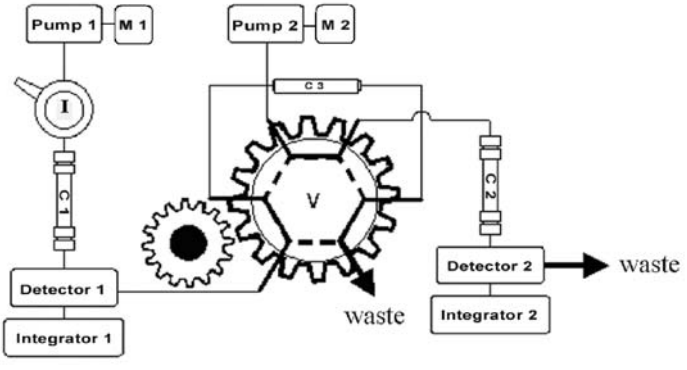


Fig. 1. A schematic diagram for the automatic column-switching HPLC system used in this study. The solid and dotted lines indicate switching valve position A and B, respectively. Abbreviations used are as follows: C1, column 1 (TSKgel ODS -80Ts); C2, column 2 (Chiralpak AD-RH); C3, column 3 (COSMOSIL® 5C18-MS- Ⅱ guard column); M1, H_2O/MeOH/CH_3CN (70/20/10); M2, H_2O/CH_3CN (40/60).

2 Experimantal

Animals

Animal experiments were approved by Laboratory Animal Research Committee of Taipei Medical University. Female C57BL/6 mice (Laboratory Animal Center of National Taiwan University, Taipei, Taiwan) were used in the experiment and kept in an environmentally controlled room with food and tap water ad libitum. To induce Nephrotoxic Serum (NTS) nephritis, C57BL/6 mice were immunized with 250 μg of rabbit IgG and 0.05 mL of Freund Complete Adjuvant (FCA) in the rear footpad. Five days later, mice were intravenously injected with 100 μL of NTS. The urine was collected using metabolic cage once a week for 12 hrs.

Samples and Derivatization procedure

Urinary D- and L-lactate were derivatized with 4-nitro-7-piperazino-2,1,3-benzoxadiazole (NBD-PZ) according to the previous study[1] as follows: 100 μL of mouse urine was added to 400 μL of CH_3CN then centrifuged, then 100 μL of the supernatant was added to 50 μL of 10 mM NBD-PZ in CH_3CN in the presence of 50 μL each of 280 mM triphenylphosphine and 2,2'-dipyridyl disulfide in CH_3CN. After standing for 3 h at 30℃, 250 μL of 0.1% trifluoroacetic acid in H_2O was added to stop the reaction. In order to remove the excess fluorescent reagent NBD-PZ, 100 μL of the resultant solution was loaded onto a mobile phase preconditioned solid-phase extraction cartridge, Empore™ SBD-RPS (4 mm/1 mL), for complete elution, another 100 μL of the mobile phase was loaded, then the elute solutions were combined and filtered with a 0.22 μm syringe filter. Twenty microliters of the elute solution was injected into the HPLC.

HPLC conditions

A column-switching HPLC system was used in the present work. As illustrated in Figure 1, the HPLC system was equipped with a AS-950 intelligent sampler (Jasco, Tokyo, Japan), two pumps (L-7100; Hitachi, Tokyo, Japan), an F-1000 and an L-7485 fluorescence detector (Hitachi), two D-2500 Chromato-Integrators (Hitachi), and a Rheodyne Model 7000 switching valve (Rheodyne, Rohnert Park, CA., USA) with a COSMOSIL® $5C_{18}$-MS-Ⅱ guard column (Nacalai Tesque, Kyoto, Japan) as a sample trap. A TSKgel ODS-80Ts column was used for isolation and quantification of the total (D+L)-lactate in mouse urine; the mobile phase was H_2O/MeOH/CH_3CN (70/20/10, v/v) at a flow rate of 0.7 mL/min. A column packed with amylose tris (3,5-dimethylphenylcarbamate) coated on silica gel (Chiralpak AD-RH; Daicel, Osaka, Japan) was used for the enantiomeric separation of D- and L-lactate eluted with H_2O/CH_3CN (40/60, v/v), and the flow rate was 0.3 mL/min. Fluorescence detection (detectors 1 and 2) was performed at 547 nm with a 491 nm excitation wavelength [1].

The integrator received the fluorescence intensity and transform it into voltage (0~1000 μv) to recorder. According to the information given from integrator, the recorder printed out the chromatograms. As the monitored peak of total lactate appeared on the chromatograms in Integrator 1, eluted total (D+L)-lactate was presumed to be isolated in the sample trap column, and the valve position must be switched from A (solid line) to B (dotted line) to introduce the total (D+L)-lactate into

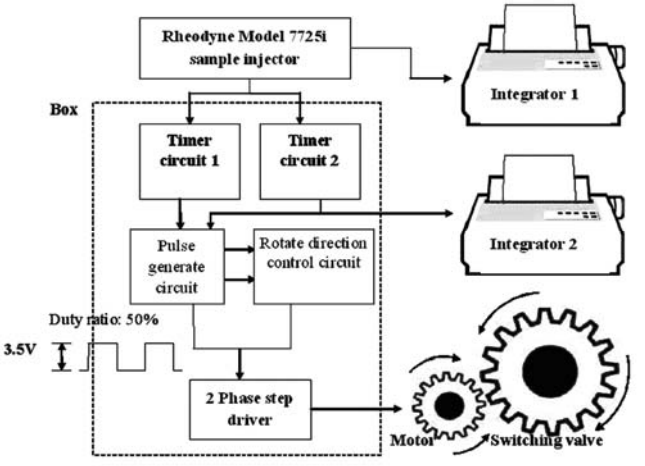


Fig. 2. The Electric hardware blocks diagram for denoting the designed relevant system

the chiral column. Then, the D and L isomers were separated enantiomerically on the chiral column, and enantioseparation chromatograms were obtained from Integrator 2.

Mechanical Design in switching valve

A two phase step motor (TS3617N3E8; TAMAGAWA, Japan) was used to rotate the switching valve in this study. As illustrated in Fig. 2, they were connected by a coupling with gear. The holding torque of step motor was 3.2 kg·cm and was operated in half step (0.9°/step), 0.2 KPPS.

Circuits

The switching valve must be switched from A (solid line) to B (dotted line) of Fig. 1 at the setting timing. Circuits can be divided into three parts: timer circuit and pulse generate circuit and rotate direction control circuit.

a. Timer circuit

Two commercial countdown circuit suites were used in this study (VCT-TDS; Tun-Hwa Electronic Material, Taiwan). Each suite mainly consists of an already programmed 89C51 to do the countdown operation, six seven-segment displays showing the time and one relay which can be used for advanced control. The relay was employed to generate a ground signal with 1 second period for triggering the pulse generate circuit. Before analyzing biological samples, we determined the retention time of racemic D, L-lactate by injection of standard. Because different chemical was isolated in different time, we designed a timer circuit to determine which compound we want to deal with.

In this case, we want to separate the total (D+L)-lactate into D and L enantiomers. According to the chromatogram of standard, we know the total (D+L)-lactate appeared around 34 min. When sample injector injected biological samples into the TSKgel ODS-80Ts column, it also produced a short signal and we took it to be the start signal

of our system. Once the timer circuits received the start signal, they began to count down. Timer 1 was set at the time when total (D+L)-lactate was going to be isolated and timer 2 was set at 4 minutes after time 1. Four minutes was the approximate time from the emergence to disappearance of the (D+L)-lactate, it might be gotten by preliminary experiments although it is different by chemicals. When timer circuit 1 counted to zero, switching valve was turned from position B to position A to introduce total (D+L)-lactate into the guard column. Four minutes later, total (D+L)-lactate was isolated completely. Timer circuit 2 would count to zero to turn the valve back from A to B and start the integrator 2 at the same time. Then the D and L enantiomers were separated enantiomerically on the chiral column. When two timer circuits counted to zero, they started pulse generate circuit to produce square waves to drive stepping motor.

b. Pulse generate circuit and Rotate direction control circuit

Pulse generate circuit was constructed mainly by a timer IC (STMicroelectronics, NE555N) and three synchronous up/down decade counters (STMicroelectronics, M74HC192). Timer IC aimed at producing square waves with 3.5V amplitude to drive the motor while counters and dip switches defined the number of waves which were required. The angle difference of switching valve between position A and B was 58˚. Due to stepping motor must switch the valve from position A to position B, the pulse generate circuit produced square waves with 50% duty ratio at 200 Hz each time to drive the motor to rotate switching valve after timer circuits count down. After each rotation, pulse generate circuit sent a signal to the rotate direction control circuit to make stepping motor always rotates in reverse direction at next time. In this way, we can also prevent the switching valve from being damaged if there was a mis-trigger driving the pulse generate circuit.

3 Application

The hardware of the automated switching valve and the HPLC system are shown in the figure 3a and figure 3b, respectively. The proposed column-switching HPLC method was applied to the urine of normal and renal failure mice. Urinary D, L-lactate

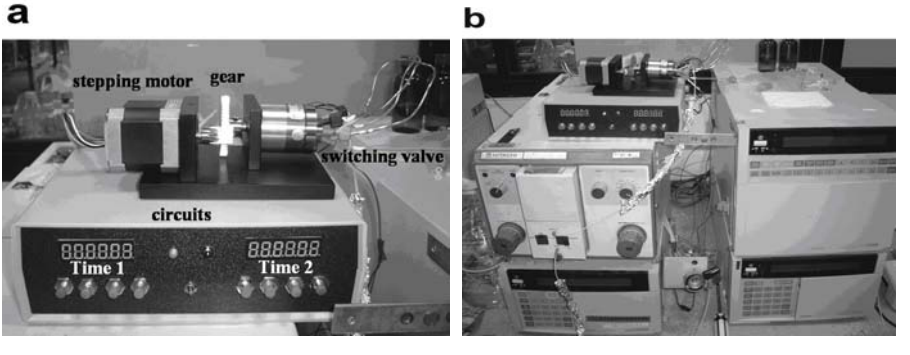


Fig. 3. (a) The hardware of the connection for gear coupling, step motor and switching valve. (b) All views of automated column-switching HPLC system.

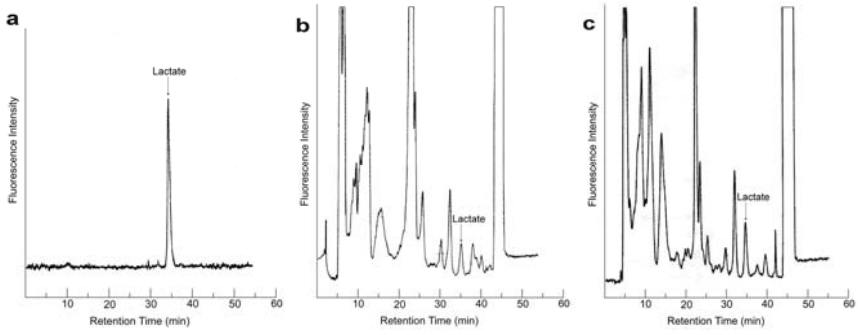


Fig. 4. HPLC chromatograms of (D+L)-lactate in normal and renal failure mouse urine, which obtained by proposed automatic column-switching HPLC system. Total (D+L)-lactate was first isolated from standard (a), normal (b) and renal failure mouse urine (c), respectively.

reacted with NBD-PZ for fluorescent derivatization, and was separated on the ODS column and determined fluorimetrically at 547 nm with 491 nm of excitation wavelength (Fig. 4a, b and c). During the separation step on the ODS column, the peak fraction of (D+L)-lactate derivatives was introduced into an amylose-type chiral column by changing the flow of the eluent via 6-port valve. Then, D-lactate derivative was separated enantiomerically from L-lactate derivative, and the enantiomeric ratio was determined from the chromatogram. Using this proposed HPLC method, 20 μL of urine sample was sufficient for D-lactate determination. As shown in Fig. 4b and 4c, urinary D-lactate showed a rising trend. Urinary D-lactate may be used as a marker to screen the stage of kidney damage.

Previously, we analyzed the samples with the valve rotated manually, and only about 10 mouse urine samples could be determined within one day. After testing our automatic column-switching HPLC system, we can successfully rotate the valve at the right time, not only in the analysis of normal mice urine, but also in the renal failure mice.

In the future, we will add the voltage slope detect circuits for making the determination of biological samples automatically. When the slope of chromatogram intensity is plus for a time ΔT, the step motor rotate. Second, we need some protect circuit to prevent the switching valve from being over-rotated and damaged.

4 Conclusion

With the automatic column-switching HPLC system, more samples could be analyzed within one day and researchers don't need to stay around the machine. It is a useful system especially when great amount of samples have to be analyzed. This automatic system makes the determination of biological samples become more convenient and may be helpful to the researchers.

Acknowledgements

The authors express their sincere appreciation to Dr. Jen-Ai Lee (Taipei Medical University, Taipei, Taiwan) for his helpful suggestions. This work was financially supported by the Educational Council of the Republic of China.

References

[1] Fukushima, T., Adachi, S., Ichihara, H., Al-Kindy, S., Imai, K.: Biomedical Chromatography 13, 418–424 (1999)
[2] Fukushima, T., Lee, J.A., Korenaga, T., Ichihara, H., Kato, M., Imai, K.: Biomedical Chromatography 15, 189–195 (2001)
[3] Ichihara, H., Fukushima, T., Imai, K.: Analytical Biochemistry 269, 379–385 (1999)
[4] Lee, J.A., Tsai, Y.C., Chen, H.Y., Wang, C.C., Chen, S.M., Fukushima, T., Imai, K.: Analytica Chimica Acta 534, 185–191 (2005)
[5] Hasegawa, H., Fukushima, T., Lee, J.A., Tsukamoto, K., Moriya, K., Ono, Y., Imai, K.: Analytical & Bioanalytical Chemistry 377, 886–891(2003)

A New Approach for Veins Detection

Dana Lodrová, Radim Dvořák, Martin Drahanský, and Filip Orság

Brno University of Technology, Faculty of Information Technology
Božetěchova 2, 61266, Brno, Czech Republic
{ilodrova,idvorak,drahan,orsag}@fit.vutbr.cz

Abstract. This paper deals with a new approach for detection of finger veins. The method is split into four steps. The first of them consists of basic series of image filtering of the vein pattern and the other three are sequences of image filters for determination of the finger contour used for the background masking.

Keywords: finger, vein, image, filtering, enhancement, infrared light.

1 Introduction

The relevance of biometric systems is increasing in both application areas, i.e. systems for criminology praxis and systems for access control. Both of them are based on different requirements: systems for criminology experts do most often the identification tasks, i.e. they try to find an offender, whereas the access control systems are oriented only on verification attempts.

We have done a long time research in the area of fingerprint recognition. However, there was an interesting question, what is beneath the skin (i.e. under the structure of ridges and valleys). In the beginning we used an infrared illumination with a classical VGA (video graphics array) camera. The results were not good, nevertheless acceptable. Hereby we started the research in the field of finger veins recognition.

There are not many systems available for finger veins recognition or scanning in the market. We have done an extensive recherché in the Internet and have found that there is only a limited number of industrial solutions offered to this topic. One of the commercially provided solutions is offered by the company Hitachi [1]; they have a longer period experience in this area. Similar technology is being offered by some other companies, e.g. Sony [2], Bioaccez Controls [3], M2SYS [4] or FDS/AABACS [5].

2 Vascular Pattern Recognition

Finger veins recognition based authentication is not a new idea. It belongs among one of the vascular pattern recognition systems. Generally, popularity of vascular pattern recognition based systems started to increase few years ago as demand after the biometric systems had started to grow due to the increased needs in the field of security.

Vein recognition is non invasive and reliable way to identification. It is well accepted by users. Devices for this sort of authentication can be small enough for a convenient every day use and can put installed at any place.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 76–80, 2009.

Systems based on the vein recognition consist typically of a capturing device and appropriate illumination. On the side of hardware the main differences lie in what is being captured and what is being illuminated and how. The software then consists of few basic image processing steps, feature extraction and decision. These three main points give us enough space for innovations and experiments.

2.1 Hardware Prototype

We designed the first prototype of a device intended for capturing of the finger veins patterns, their processing and recognition/verification. The device consists of an illumination unit with infrared diodes (recommended wavelengths are around 900 nm), a digital signal processor for video pre-processing, image enhancement and processing and adjustment of the diodes. Then, there is a microcontroller to control the peripherals and USB host connection. The last part of the device is a memory to store the enciphered templates. The housing of the prototype is made of aluminum whereas the final version will be made of plastic. The device is connected and powered via a USB port.

2.2 Algorithm

Camera in the first prototype of our sensor provides us a 640×480 pixels image. The image has to be cropped so that there is remaining only a small area (119×354 pixels) containing the image of the finger. This crop is only a basic and non-precise step with a big tolerance. A precise position of the finger is determined in a later stage of the image processing. We use this preliminary step to decrease memory requirements, because the process is intended to run in the sensor due to the security reasons.

The crop is an input image to a sequence of image filters (see Fig. 1). After application of the sequence of the filters there is a structure/skeleton of the finger veins remaining and the position of the finger is marked (left side, right side and top of the finger). The detected position of the finger is used for a more precise crop of the image and for masking of the background noise around the finger. Used filters and their parameters had to be very carefully chosen and put together, because even a small change in size of some of them could propagate itself through the sequence of the filters and could cause much bigger errors and, therefore, inapplicableness of the extracted skeleton.

2.3 Thinning of the Veins Pattern

First, we describe the sequence of filters used for veins detection. This sequence consists of five functional blocks – image filters.

The first block enhances specific features of the image. It consists of a median filter (size of 5×5 pixels) and a common smooth filter (size of 3×3 pixels). Size of the individual filters was determined experimentally.

The second block performs a convolution. In this case, we use a special convolution kernel, which causes an effect similar to a side illumination of the finger (see Fig. 1b). This effect enhances relief of the finger veins.

After disclosure of the relief of the veins, it is necessary to process the image again, which is purpose of the third block of filters. First, a binarization is performed

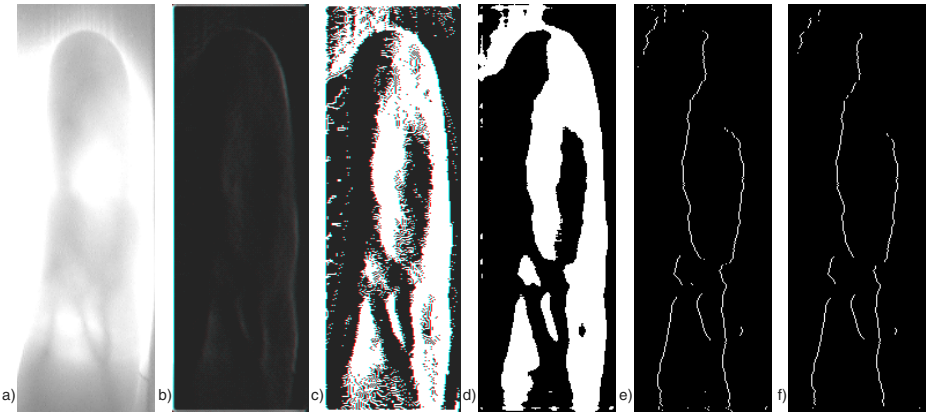


Fig. 1. A sequence of processed images: a) original image, b) image after applying a convolution, c) threshold image, d) image enhanced by a median filter, e) extracted vein skeleton, f) filtered image by the "special median" without the noise pixels.

with a threshold T=0 (so that values greater than 0 are converted to 255). Then, small inaccuracies are fixed by a median filter (size of 7×7 pixels).

The fourth block consists of a special thinning algorithm. The algorithm is simple but very efficient. It just goes through all image rows and looks for a special pattern (three black pixels followed by three white pixels). If this sequence of pixels is found, then a white pixel is written in an initially black output image. The result of this procedure is one pixel thin skeleton of finger veins.

Last block of the image filters consists of so called "special median" filter, which is a special filter with a kernel of size of 3×3 pixels removing all undesirable isolated points (the noise on the background) from the image.

2.4 Detection of the Top of a Finger

To create a template it is important to locate the finger – hence, to determine its position. One important part of the finger useful for the localization of the finger is the top of it. Position of the top provides us all necessary information about the finger position and orientation since the finger is more or less fixed in a vertical position. The only one direction the finger could be shifting is the vertical one because different people have fingers of various lengths or they do not put the whole finger into the camera view, which can cause vertical shift of the finger.

The top of the finger is detected simply as the first occurrence of a horizontal line in the image, which is reached in the stage of the image preprocessing done in the basic process described before. Namely, it is the part of the image smoothing by median filtering with the kernel size of 5×5 pixels. The smooth with the kernel of size 3×3 pixels is the second filter that is applied afterwards.

The half of the Sobel edge detector is applied after the image preprocessing. The horizontal version of the Sobel kernel was chosen. Result of application of the sequence of filters is shown in Fig. 2a.

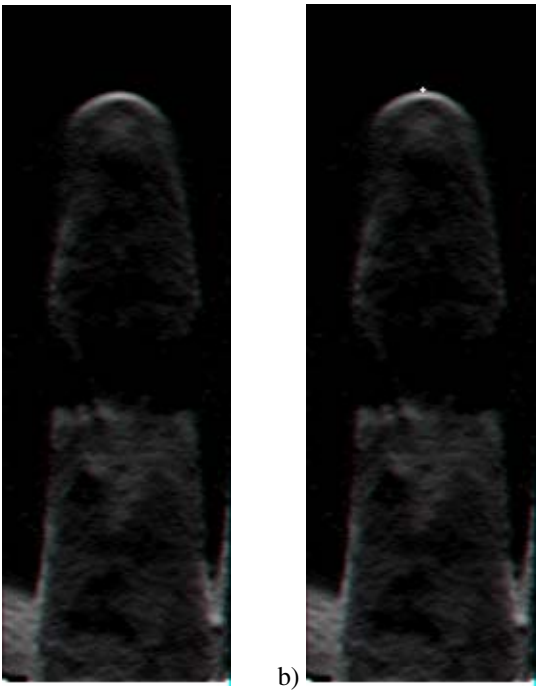


Fig. 2. a) Image after application of the filter; b) Image with the detected finger top

The top of the finger can be then easily found as it is the first occurrence of the white pixels or sequence of white pixels in a row. In the current version we use the sequence of three points in one row that are brighter than the selected threshold. The example of the detected top of the finger is shown in Fig. 2b.

2.5 Detection of the Finger Contour

Additional information that can be included into the finger vein template is a contour of the finger. It can help to erase false detected veins in the image lying outside of the finger. The incorrectness is caused mainly by the lighting condition.

During the contour detection we are trying to use results of the previously used sequence of filters as much as possible in order to reduce the overall computational cost. Hence, we are trying to use already computed values. We are detecting the left and the right part of the contour separately beginning from the top of the finger.

During the detection of the finger contour the errors arise due to the noise and the overall quality of the image. The errors are of a form of shifted pixels or group of pixels in a horizontal direction. If we assume that the most part of the detected contour is correct (in the all tested cases the bigger part of contour was detected correctly) and if we take into account that the contours are continuous, then we can relatively easily make the contour correction.

3 Conclusion

In the future we are planning to perform much more detailed tests with hundreds of volunteers. We proposed a method for quality setting and we will apply it to the database of templates in order to improve accuracy of the device.

We are also planning several enhancements of the described algorithms. The first of them is reduction of the template size in order to lower the computational cost and memory requirements. The next improvement is the speedup of database searching. The basic idea of this improvement is based on the division of the templates into few specific groups with similar vein structures.

With respect to the future developments of the algorithm, we are preparing enhancements of the vein detection such as a more sophisticated correction of the veins. The correction will be based on reduction of the false detected veins and on the connection of the proper ones.

The proposed device for finger vein detection will be soon applied for patent at the Czech Patent Agency.

Acknowledgements. This research has been done under support of the following three grants: "*Security-Oriented Research in Information Technology*", MSM0021630528 (CZ), "*Information Technology in Biomedical Engineering*", GA102/09/ H083, "*Education of Liveness Testing in Subject Biometric Systems*", FR2525/2009/G1 and by companies *BetaLab s.r.o.* and *Digitus s.r.o.*

References

1. Hitachi Finger Vein Authentication Technology (2009), http://www.hitachi-america.us/products/business/smart_solutions/finger_vein/ (cit. 2009-07-25)
2. Sony morfiria finger-vein biometric system (2009), http://www.slashgear.com/sony-mofiria-finger-vein-biometric-system-announced-0232716/ (cit. 2009-07-25)
3. Bioaccez Controls Bio-Veinz finger vein reader (2009), http://www.bioaccez.com/en/articles/22/3/biometrics-products/Page3.html (cit. 2009-07-25)
4. M2SYS M2-FV finger vein reader (2009), http://www.m2sys.com/finger-vein-reader.htm (cit. 2009-07-25)
5. Aabacs finger vein authentication reader (2009), http://aabacs.com/main/p_ai200.php (cit. 2009-07-25)

An Analysis of Social Guarantees for Context Based Applications*

Juanita Pedraza[1], Miguel A. Patricio[2], Agustin De Asís[1], and Jose M. Molina[2]

[1] Public Law Department
[2] Computer Science Department
Universidad Carlos III de Madrid, Colmenarejo, Spain
{jpedraza,aeasis}@der-pu.uc3m.es, mpatrici@inf.uc3m.es,
molina@ia.uc3m.es

Abstract. The biometric identification in context based applications is a promising research area but several legal aspect should be taken into account to develop context applications following legal assumptions as privacy, human rights, etc.. In this work, we present a guide of social guarantees to be developed in the context system for private relations (private user and private services) as for public relations (private/public users and public services). The proposal of this paper is center in a set of goals that should be accomplished by context applications to guarantee the privacy and the human rights.

Keywords: Context Applications, Biometrics Identification, Social Guarantees, Privacy and Human Rights.

1 Introduction

The development of systems based on the concept "context-aware computing" has been an intention to transform the way people interact with new technologies (every time smaller and smarter). Context aware computing was firstly defined by Shilit [1] and he claimed that the main components of context were nearby. He was referring to: who you are, what are you doing, where you are, when and why. There is also a more widely accepted and used definition of what context is, and it was given by Dey [2] where he defines context as: "any information that characterizes a situation related to the interaction between humans, applications, and the surrounding environment."

There are several developing systems such as platforms, frameworks and applications for offering context-aware services, inter alia, the Context Toolkit proposed in [2] Context Fusion Networks [3] and Context Fabric [4]. A recent one is Appear which is a context-aware platform designed to provide contextual information to users in particular and well defined domains. It has a modular architecture and we have already used it in a previous work [5].

* This work was supported in part by Projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, SINPROB, CAM MADRINET S-0505/TIC/0255 and DPS2008-07029-C02-02.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 81–88, 2009.

In Europe, the concept of Ambient Intelligent (AmI) includes the contextual information but expand this concept to the ambient surrounding the people. So, electronic or digital part of the ambience (devices) will often need to act intelligently on behalf of people. It is also associated to a society based on unobtrusive, often invisible interactions amongst people and computer-based services taking place in a global computing environment. Context and context-awareness are central issues to ambient intelligence [6]. AmI has also been recognized as a promising approach to tackle the problems in the domain of Assisted Living [7]. Ambient Assisted Living (AAL) born as an initiative from the European Union to emphasize the importance of addressing the needs of the ageing European population, which is growing every year as [8]. The program intends to extend the time the elderly can live in their home environment by increasing the autonomy of people and assisting them in carrying out their daily activities.

Several prototypes encompass the functionalities mentioned above: Rentto et al. [9], in the Wireless Wellness Monitor project, have developed a prototype of a smart home that integrates the context information from health monitoring devices and the information from the home appliances. Becker et al. [10] describe the amiCa project which supports monitoring of daily liquid and food intakes, location tracking and fall detection. The PAUL (Personal Assistant Unit for Living) system from University of Kaiserslautern [11] collects signals from motion detectors, wall switches or body signals, and interprets them to assist the user in his daily life but also to monitor his health condition and to safeguard him. There are also several approaches with a distributed architecture like AMADE [12] that integrates an alert management system as well as automated identification, location and movement control systems.

All these approaches are promising applications from an engineering point of view, but, no legal aspects are considered in the development. Clearly, an important point is the necessity to identify the users of these systems. Two different approaches could be considered, one approach is based in the cooperation of the user to be identified and another one is based in the non-cooperative environment (for example in surveillance applications).

In this work, authors define a set of procedures that should be contained in the context-aware applications to accomplish the legal aspect in Europe and USA related to privacy and human rights. Section 2 is center in biometric identification techniques, analyzing several alternatives and the necessity of cooperation by user. In section 3 a general review of the legal normative in Europe and USA is described to show the main points to take into account. A description to the necessities to be included in context-aware applications is enumerated in section 4. Finally, in section 5 some conclusions are included.

2 Context Aware Applications and Biometric Identification

Identification and personalization are essential features of context-based services. The development of efficient, non-vulnerable and non-intrusive biometric recognition techniques is still an open issue in the biometrics field (in which, however, enormous scientific contributions have been made over the last decade). It is also a necessity to obtain contextual systems able to provide a satisfactory user experience.

Table 1. Comparison of several Biometric Identification Procedures

Biometric Technique	Verify	Identify	False Positive	False Negative	Intrusiveness	Cost
Face recognition (2D)	Yes	No	Hard	Easy	Very Low	Low
Fingerprint	Yes	Yes	Very Hard	Very Hard	Middle	Low
Hand geometry	Yes	No	Very Hard	Middle	Low	Middle
Iris Scanning	Yes	Yes	Very Hard	Very Hard	Middle	High
Retinal Scanning	Yes	Yes	Very Hard	Very Hard	High	High
Voice Recognition	Somes	No	Middle	Easy	Very Low	Low
Signature	Some	No	Middle	Easy	Low	Middle

Reliable biometric system has long been an attractive goal. The Biometric identification must be robust, efficient and quick process to be accepted in strong requirements of security in this networked society [13]. Biometrics aims to recognize a person through the physiological or behavioral attributes [14], such as iris, retina, fingerprints, DNA and so on. The cause for this increment in research fields is the security sector and the possible application in many its aspects, such as video-surveillance or access control. A summary of identification procedures is presented in table 1, where the classical concepts of verification, identification, false positive, false negative, intrusion and cost are compared among several classical biometric techniques.

The contextual framework needs a biometric scheme with the following features:

- Multibiometric: which combines several sources of biometric information (traits, sensors, etc.) with the aim of mitigating the inherent limitations of each source, obtaining a more reliable and accurate system.
- Highly transparent, highly accepted, and low intrusive, using biometric traits that can be acquired even without any cooperation of the user (e.g. face, voice) and well socially accepted (like the handwritten signature)
- Able of inferring human activity and analyzing user emotions, therefore significantly focused on services customization.

These requirements affect directly to many legal aspects that should the considered before the development of industrial applications, to be used in the private sector or public sector.

3 Legal Aspects in Biometric Identification

Biometric technology has legal implications because it has the potential to reveal much more about a person than just their identity. For instance, retina scans, and other

methods, can reveal medical conditions. Thus biometric technology can be a potential threaten to privacy [15].

European and American judges [16] have categorized privacy as taking three distinct forms. These includes [17]: a) physical privacy or freedom from contact with other people; b) decisional privacy or the freedom of the individual to make private choices about the personal and intimate matters that affect her without undue government interference and c) informational privacy or freedom of individual to limit access to certain personal information about oneself. Obviously, biometrical technology is related with the a) and c) issues.

Biometric identification, of course, is not a new technology. Introduced more than a century ago, fingerprint technology is perhaps the most common biometric identification technique. Thus the social risk [18] associated to this technology is not new. However, technological advances, among other factors [19], have increased the social risk associated to technique because: a) they have reduced the social tendency to reject its use; b) they have allowed their widespread use [20] and c) they have enabled to obtain more sensitive information on the subject.

States and stakeholders should make further efforts to ensure that biometrical applications are monitored and the rights and freedoms of individuals are respected [21]. In particular, they should take into account, inter alia: the legal nature of relations (public or private) and the characteristics of the devices (ability to obtain sensitive information):

a) Private Relations (Private Users and Private Services) [22]

Because most biometric scanning will result from private sector activities where the user voluntarily gives up information, legal privacy concerns will usually be implicated to ensure informed consent and the transparency with the data subject. This is achieved providing them with the information about the systems and granting the right to access to personal data and, where appropriate, the right to have it deleted or rectified or blocked if they are inaccurate or have been unlawfully processed [23].

b) Public Relations (Private/Public Users and Public Services)

In this context, the social guarantees, depends on the particular case and the results of legal test of the "balancing interests" [24][25]. There are common principles to "balancing interest" test: proportionality and reasonableness.

The principle of proportionality requires that measures implemented should be appropriate for attaining the objective pursued and must not go beyond what is necessary to achieve it. The reasonableness of a measure is therefore to be adjudged in the light of the nature and legal consequences of the relevant remedy and of the relevant rights and interests of all the persons concerned.

Also in this field, States shall ensure that appropriate procedures guaranteeing the dignity and privacy of the applicant, in particular, the protection of personal data. The States concerned shall closely monitor the implementation of the social guarantees, including: a) the general information on features and uses of systems; b) all the technical and organisational security measures required to protect personal data

against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, and all other unlawful forms of processing the personal data; c) the collection and transmission of biometric identifiers; d) any processing of personal data must be lawful and fair to the individuals concerned; whereas, in particular, the data must be adequate, relevant and not excessive in relation to the purposes for which they are processed; whereas such purposes must be explicit and legitimate and must be determined at the time of collection of the data; whereas the purposes of processing further to collection shall not be incompatible with the purposes as they were originally specified; e) in all cases the level of security shall be adapted to the sensitive nature of the data; f) in general, the techniques taken to ensure compliance with data protection provisions and provide a mechanism for citizens to access, control, and verify their information.

Society as a whole needs to be aware of the obligations and rights that are applicable in relation to the use of biometric applications. Therefore it makes sense to create a regulatory model for the collection, use and dissemination of biometric information. In that regard, there're several options like laissez faire approach, self-regulation, public regulation [26]. Under a laissez faire regime, no authority requires businesses to disclose their biometric policies to consumers. Therefore, it would be difficult for customers to comprehensively weigh the alternatives. The self regulation is not sufficient because entails one big drawback: the lack of enforcement. The last alternative deals with binding legislation with effective, proportionate and dissuasive sanctions for infringements.

4 Development of Context Applications with Social Guarantees

As a complementary way for a public regulation, the software development industry, by proper or government initiative, could introduce in the design of the context based applications some social guarantees, commented in the previous section. These social guarantees should be introduce, specially, when non-cooperative biometric identification is carried out, in order to preserve the privacy and human rights of the user. Many legal requirements could be easily introduce in the context based applications if they are considered in the analysis and design phase.

From the point of view of the development of this kind of context applications, all systems should satisfy the requirements of the Tables 2 and 3. As in a software engineering project, each requirement has an identifier, a description of the requirement, need, priority and stability. The need is to indicate whether or not the essential condition to which it refers. Stability indicates whether the requirement can be modified or not. As a classical user requirement document, we have described three types of priorities: priority 1 (the system must have this requirement); priority 2: (the system should have this requirement) and priority 3 (the system could have this requirement).

Where are the facilities to be included in context based system to be used between user and companies? In this case, applications should consider the requirements of Table 2.

Table 2. User Requirements in context based system to be used between user and companies

ID	Description	Need	Prio.	Stability
URC01	The user must know the technical characteristics, the results of the application and the use that company could do to these results.	Yes	1	Stable
URC02	The user must give express consent.	Yes	1	Stable
URC03	The user could revocation, in any time, his express consent.	Yes	1	Stable
URC04	The user should be able to execute the whole rights integrated in the legal systems of personal data protection [27].	Yes	1	Stable
URC05	The data stored and the processed one should accomplish the legal security norms.	Yes	1	Stable

Where are the facilities to be included in context based system to be used between user and government applications? In this case, applications should consider additional requirements depicted in Table 3.

Table 3. Additional User Requirements in context based system to be used between user and government applications

ID	Description	Need	Prio.	Stability
URG01	The development application should be inside the general principles of reasonableness and proportionality.	Yes	1	Stable
URG02	Equal to URC01	Yes	1	Stable
URG03	The application should respect, in any case, the user personal dignity.	Yes	1	Stable
URG04	Equal to URC04	Yes	1	Stable
URG05	Equal to URC05	Yes	1	Stable

Finally, the system for social guarantees needs the verification of the compliance with these requirements, independently of the origin of these requirements: public norms or auto regulative process. These requirements should be verified in a previous process before the commercialization or distribution, as a software/hardware license or authorization. The liability should include to the developer, the distributer and the user. This subjective scope of liability, joint to sanctions (criminal or administrative), constitutes the final closure of a legal system for monitoring compliance with social guarantees demanded by the widespread use of these techniques.

5 Conclusions

In this paper, we present the necessity to consider legal aspect, related with privacy or human rights, into the development of the incipient context based services. Clearly,

context based services and Ambient Intelligence (and the most promising work area in Europe that is Ambient Assisted Living, ALL) needs a great effort in research new identification procedures. These new procedures should be non-intrusive, non-cooperative, in order to the user be immersed in a Intelligent Environment that knows who is, where is and his/her preferences. These new paradigms should be development accomplished the legal issues to allow users be citizen maintaining their legal rights.

References

1. Shilit, B.N.: A Context-Aware System Architecture for Mobile Distributed Computing. Ph.D.thesis, Dept of Computer Science, Columbia University (1995)
2. Dey, A.K., Saber, D., Abowd, G.D.: A conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Human-Computer Interaction (HCI) Journal 16, 97–166 (2001)
3. Chen, Guanling, Kotz, David.: Context Aggregation and Dissemination in Ubiquitous Computing Systems. In: Proceedings of the Fourth IEEE Workshop on Mobile Computing Systems and Applications, June 20-21, 2002, p. 105 (2002)
4. Hong, J.: The context fabric: An infrastructure for context-aware computing. In: Minneapolis, A.P. (ed.) Extended Abstracts of ACM Conference on Human Factors in Computing Systems (CHI 2002), pp. 554–555. ACM Press, Minneapolis (2002)
5. Sanchez-Pi, N., Fuentes, V., Carbo, J., Molina, J.: Knowledge-based system to define context in commercial applications. In: Proceedings of 8th International Conference on Software Engineering, Artificial Intelligence, Networking, and Paraallel/Distributed Computing (SNPD), Qingdao, China (2007)
6. Schmidt, A.: Interactive context-aware systems interacting with ambient intelligence. IOS Press, Amsterdam (2005)
7. Emiliani, P., Stephanidis, C.: Universal access to ambient intelligence environments: Opportunities and challenges for people with disabilities. IBM Systems Journal 44(3), 605–619 (2005)
8. World population prospects: The 2006 revision and world urbanization prospects: The, revision. Technical report, Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat (last access: Saturday, February 28, 2009; 12:01:46 AM)
9. Rentto, K., Korhonen, I., Vaatanen, A., Pekkarinen, L., Tuomisto, T., Cluitmans, L., Lappalainen, R.: Users' preferences for ubiquitous computing applications at home. In: First European Symposium on Ambient Intelligence 2003, Veldhoven, The Netherlands (2003)
10. Becker, M., Werkman, E., Anastasopoulos, M., Kleinberger, T.: Approaching ambient intelligent home care system. In: Pervasive Health Conference and Workshops 2006, pp. 1–10 (2006)
11. Floeck, M., Litz, L.: Integration of home automation technology into an assisted living concept. Assisted Living Systems-Models, Architectures and Engineering Approaches (2007)
12. Fraile, J., Bajo, J., Corchado, J.: Amade: Developing a multi-agent architecture for home care environments. In: 7th Ibero-American Workshop in Multi-Agent Systems (2008)
13. Jain, A.K., Bolle, R.M., Pankanti, S.: Biometrics: Personal Identification in a Net-worked Society. Kluwer, Norwell (1999)

14. Daugman, J.: Biometric Decision Landscape,Technique Report No. TR482, University of Cambridge Computer Laboratory (1999)
15. That right is enshrined in Article 12 of Universal Declaration of Human Rights, Article 7 the Charter of Fundamental Rights of the European Union (2000/C 364/01) and implicity in Fourth Amendment
16. See: European Court of Human Rights, López Ostra v. Spain - 16798/90 [1994] ECHR 46 (9 December 1994). Katz v. United States, 389 U.S 347 (1967) Skinner v. Railway Labor Executives' Ass'n, 489 U.S. 602 (1989). To see differences between legal systems: Kirtley: Is implementing the EU Data Protection Directive in the United States irreconcilable with the First Amendment? In: Government Information Quaterly, vol. 16(2), pp. 87–91 (2001)
17. Woodward, J.: Biometric scanning, law & policy: identifying the concerns-drafting the biometric blueprint. U. Pitt. L. Rev. 59, 97–155 (1998)
18. Beck, U.: La sociedad del riesgo: hacia una nueva modernidad (1998)
19. Lin, Liou, Wu: Opportunities and challenges created by terrorism. Technological Forecasting and Social Change 74(2), 148–164, 158 (2007)
20. Kennedy, G.: Thumbs up for biometric authentication. Computer Law Review & Tech. (8), 379–407 (2003-2004)
21. Parejo Alfonso, L.: Seguridad pública y policía administrativa de seguridad, Valencia (2008)
22. To see examples, `http://www.biometrics.gov/Documents/FAQ.pdf` (040809)
23. That rights are enshrined in Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data OJ L 281, 23.11.1995, p. 31–50 and Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector OJ L 201, 31.7.2002, p. 37–47. In the United States doesn't exist general regulation for data protection. Kuner, C: An international legal framework for data protection: issues and prospects. Computer Law & Security Review 25, 307–317 (2009)
24. Haas, E.: Back to the future? The use of biometrics, its impact on airport security, and how this technology should be governed. Journal of Air Law and Commerce (69), 459 (spring 2004)
25. Rodríguez de Santiago, J.M^{a}.: La ponderación de bienes e intereses en el Derecho Administrativo. Madrid (2000)
26. Kennedy. Note 20
27. This type of regulatory systems has been used previously in access points in the frontiers of the European Union, where a specific regulation exists: Council Regulation (EC) No 2252/2004 of 13 December 2004 on standards for security features and biometrics in passports and travel documents issued by Member States OJ L 385, 29.12.2004, p. 1–6 modified by Regulation (EC) No 444/2009 of the European Parliament and of the Council of 28 May 2009 amending Council Regulation (EC) No 2252/2004 on standards for security features and biometrics in passports and travel documents issued by Member States OJ L 142, 6.6.2009, pp. 1–4

Data Warehouse Approach to Build a Decision-Support Platform for Orthopedics Based on Clinical and Academic Requirements

Sheng-Hui Lin[1,2], Yuan-Chii G. Lee[2], and Chien-Yeh Hsu[2]

[1] Orthopedic department, CHIMEI medical center,
No. 901, Zhong Hua Rd, Yongkang City, Tainan County, Taiwan
dantes@mail.chimei.org.tw
[2] Graduate Institute of Medical Informatics, Taipei Medical University,
No. 250 Wu-Xing Street, Taipei City, Taiwan

Abstract. The continuous quality improvement has become a trend in the contemporary medical society, and that can be achieved by the specialty database implement. Decision-support system in the academic and clinical aspects are included in the process such continuous quality improvement. The database has its limitation in the decision-support due to deficiency of on-line analytic function. The data warehouse offers the sophisticated function for decision-support processes. However, the implement of data warehouse may face a lot of obstacles, included expensive cost and large personnel. We had previously established a database of orthopedics, which collected the patients' data since 2002. The new system was constructed based on this specialty database, the knowledge architectures was build up via specialists committee and accreditation indicators. The major function was to generate sufficient information for decision-support process in the academic and clinical aspects. The execution efficiency of this system is more effective than database. The unique knowledge architecture can form a distinguishing feature of the department. The cost that saved from personnel and time reduced from reports generation for accreditation is remarkable. The stratification of web-based interface application can be assessed through questionnaires; the outcome is satisfactory as what we previously expected. The sophisticate function of the data warehouse is hard to express in a solitary department of the hospital, especially when they had already owned traditional database. The experience of this system construction can be useful as one option for upgrade of specialty database and a step forward to the goal of the continuous quality improvement.

Keywords: computerized, orthopedics, HIS, database, data warehouse.

1 Introduction

The importance of the specialty database -- As the purpose of computer systems for patient data management, the comprehensive information system became the essential component in clinical practices. A lot of specific systems had been developed and published, such as COSTAR by McManus[1] and OASYS by Stoodley[2]. The former

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 89–96, 2009.

COSTAR is an integrated, modular system that can be implemented incrementally for medical records, accounts receivable, scheduling, and report generation. The latter OASYS is a kind of medical auditing system based on individual diagnoses or specific therapies.

In the clinical orthopedic practice, enormous patient amount and fast admission turnover result in overloading of medical staffs. Decision-making process in such practice needs abundant information from variable data sources. In order to obtain significant information for decision-making process efficiently, organized data are crucial. The organized data have some characteristics include objectivity, accuracy and analyzable.

Advantage and disadvantage of data warehouse -- Early in 1994, Ruffin[3] had raised the importance of medical application of data warehouse. Prather[4] had published a paper in annual meeting of American Medical Informatics Association (AMIA) on 1997, which described new medical knowledge could be generated via data warehouse discovering. The medical institutes and groups implement the data warehouse increasingly in the recent year. They had developed several experiences and techniques to implement their data warehouses. Kerkri[5] had developed their data warehouse focus on the epidemiologic domain and named this system as EPIDWARE. EPIDWARE is an integrated system for providing access to a collection of heterogeneous medical information systems.

Data warehouse can make up the deficiency of databases. There are commercial products of data warehouse in the market. Implement of such systems is seldom found in the solitary department, because of their huge scale and expensive cost. The professional information technician is necessary for maintaining the system integrity. Those are disadvantages of data warehouse applied in the orthopedic department.

Establishment the unique system -- We create a unique system for clinical and academic use of orthopedic departments, which possess functions of auto-extraction of patient data source from the hospital information system (HIS) and on-line analytic process (OLAP) of the data-warehouse. Through this system, we can obtain valuable information to perform the decision-support process and collections of specific patients' information in HIS.

2 Materials and Methods

2.1 Clinical Materials and Physical System Setup

This system serves as decision-support system that contains essential variations for decision-making process. The evaluation of efficiency and power of decision-support process in this system was designed by system executing analysis and questionnaire of users. That will be described in the following paragraphs. This system enrolled the orthopedic patients' information of a medical center in Tainan since 2002. The patients' information mainly contains admission, operation and discharge electronic healthy records. At the mean time, the overall amount of database was about 36,000 records and was increasing by days.

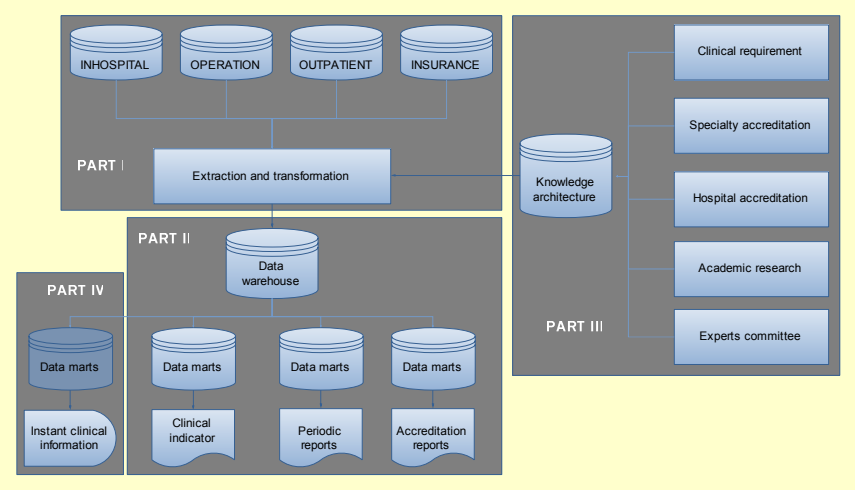


Fig. 1. Diagram of system architecture

The system is installed in a computer, which equipped with Intel® Pentium 4 and 2 gigabyte RAM. The operation system is Microsoft® WINDOWS XP professional edition. This computer is set up as a server, which is located in the intranet of the hospital. The complete set of HIS executing programs are installed in this server. The communication between HIS and the system is via intranet. The end user can access this system through any clients in the hospital.

2.2 System Architecture

Part I of the system -- The system can extract the daily records from HIS every 5 minutes by a Visual Basic (VB) based subroutine and verifies their consistence to pre-existed data. Through this process, we can ensure the timeliness and accuracy of the database. In the data extraction process we encounter some situations of data error, which included missed entries and conflicted data. Delayed data input and duplicated input by different persons may be the main causes. The medical staff may input the data on the following days after admission, or the operation records may be updated after several days of the admission. Because the data extraction from HIS is occurred on admission date, there was no update can be done in this system while records had been changed beyond the admission date. To correct these invalid data, we design an automatic periodic procedure to extract recent two months data retrospectively every 2 hours. Part II of the system -- The main framework was built by Access®, which possesses four different components – table, form, query and report. The system has three main tables of data that include inpatient, surgery and outpatient source. The tables represent the raw data extracted from HIS, which comprise all fields mapping in databases of HIS in addition to some fields of coding systems. There are several different forms for monitoring and manipulation of data by administrator. Part III of the system -- The knowledge architecture was formed in two fashions, one is individual requirements and the other is the conclusions of expert committee.

We gathered clinical questions from physician's demand. For example, what is mortality rate in the geriatric hip fracture population of the medical center? The expert committee was hold monthly; it discussed issues and indicators about the research and surgeries of the department. Such information can serve as topics of academic researches and department accreditation data. The questions from individual physician and research topics can be formed as queries of database. The indicative surgeries statistics can be created by specific queries and calculations. Fundamentals of such productions are SQL queries. To form a "WHERE" clause will test the queries results from several times by subspecialty physicians. The knowledge architecture forming process is a continuing task in our department, which was evolved after every expert meeting and new questions, developed in daily practices. Part IV of the system –- The final query outcome that includes groups of patients' data and statistics results are demonstrated through web interface. Our physicians can access this web interface via intranet of our hospital. Every physician has their own authority to look up their patient collection and specific patients populations of their researches. The system will generate the statistic reports periodically. Most of web interface are written by ASP and ASP.net. The multiple web pages are connected to a single database located in the web files directory. This database comprises user-requesting data stored in a fragmented fashion.

2.3 System Evaluation

System evaluation was divided into two aspects – actual executing efficiency and Statistics of invalid data.

Actual executing efficiency is mainly focus on the difference of traditional database query executing and the on-line analytic process of this system. We calculate the executing time of several queries and access error times in the traditional HIS database. The same calculations were applied on this system to reveal their differences.

The invalid data are defined as inconsistent records, missing records and typing error. Such invalid data can not be eliminated by retrospective data extraction. They already exist in the HIS. We adopt specific SQL queries and data mining techniques to generate the periodic report of invalid data and removed them from the system. We compare the statistics before invalid data elimination with that after elimination process to reveal the significant error amount of HIS.

3 Results

3.1 Production

The production of the system can be presented as reports and forms. The former can be generated by manual or automatic operations. The latter is created as the web-form interface which possessed interaction function and reveal the immediate requested information. The reports of the system can be divided into five aspects: accreditation reports, periodic reports, academic resources and indicators.

Accreditation report -- Accreditations in the orthopedic department include hospital and specialty accreditation. The core values of hospital accreditation are to provide a safe, effective, patient-centered, timely, efficient and equal healthcare, to establish

teamwork to meet community's health need and encouraging hospitals to pursue excellence in their own specialties. The system can generated statistics reports of indicator surgeries through manual or automatic fashion, moreover, the number and type of indicator surgeries also can be appended and changed. The indicative surgeries are defined through experts committee, accumulative surgeries amount are presented by pivot table queries, which source is obtained through complex SQL query. Because of single query condition can not encompass the whole indicative surgery; we use multiple filed query and text mining technique to obtain the accurate clinical information.

Periodic report -- The general population of orthopedic patients is enormous and turnover rate is fast. To obtain these information are difficult through HIS, which is not designed for statistics purpose. After mastering this information, the director of department can handle the change of information rapidly.

Academic resource -- The clinical studies are mainly relied on clinical information of patients. This system can support study proceeding in prospective and retrospective fashion. In the retrospective study, we receive the subjects from researchers and reform their requests to query of Access®. The SQL WHERE clauses are complex and rewritten for several times until output results can meet researchers' requirements. In the prospective study, queries are formatted as the similar form as that formatted in the retrospective study. They can monitor the two groups of patients in the comparison study; do the comparison of pre-defined parameters and accumulation the patient population.

Indicator -- Since 1999, Taiwan Joint Commission on Hospital Accreditation had introduced the USA's International Quality Indicator Project (IQIP) and developed Taiwan Quality Indicator Project (TQIP). The Commission has been dedicated to promoting quality indicator projects, provided applicable data and information for domestic hospitals. TQIP includes many indicators of different aspects. Manual calculation for indicators is laborious and time-consuming. The easier way to obtain the indicator is extraction of specific data from HIS. The system also offers such functions, which include unplanned/unexpected re-admissions within 14 days, unplanned return to the operating room during the same admission and surgical complication.

Form -- The web interface can provide functions of immediate access and custom query of clinical information. Application Service Provider (ASP) serve as an easy application tool in the internet use. That faces some problems in the direct connection to the system. Because the system has periodic updating process from HIS, therefore, the end-user's application of ASP will be block by this updating process. That will cause great inconvenient to the end-user. Our solution is to create another database in the directory of web pages. This database only encompasses tables and queries that contained individual physicians' data (not entire department data) and are created according to end-user requests. In the setting period, the system will update this remote database. Through this method, main system processing seldom interrupts web interface usage by end-users.

3.2 Efficiency

Executing efficiency -- In the relational database, time spent in a query will increase as numbers of records increase. (Comparison table each year increment) Generally,

time spent in a simple query will not disturb the use of the system possessed orthopedic patients' information of at lease eight years. Pivot table queries (especially multiple query conditions in single column) will spend a lot of time; and affect the whole system execution. Most of results of pivot table queries encompass yearly statistics data. We use the remote database to store results of pivot queries and can be revealed very fast when a query action executes. Results of pivot queries are updated every five hours. That makes a balance between easy of use and accuracy of data. What we compare are mainly in the indicative surgeries. The executing efficiency comparison in time consumption between local machine and remote client is shown in Figure 2.

Invalid data detection -- In the system, we make four queries to detect invalid data, which include wrong physician (not in the orthopedic department), wrong patients' basic information, incomplete operative records and incomplete admissive records. All that invalid data will result in incorrect statistics and reflect some problems in HIS programming and medical behavior. In errors of patients' basic information statistics, the decreasing number reveals improvement of HIS programming or debugging. The incomplete medical records (admissive and operative records) reflect medical staffs can not finish their task in time; may cause some legal problems.

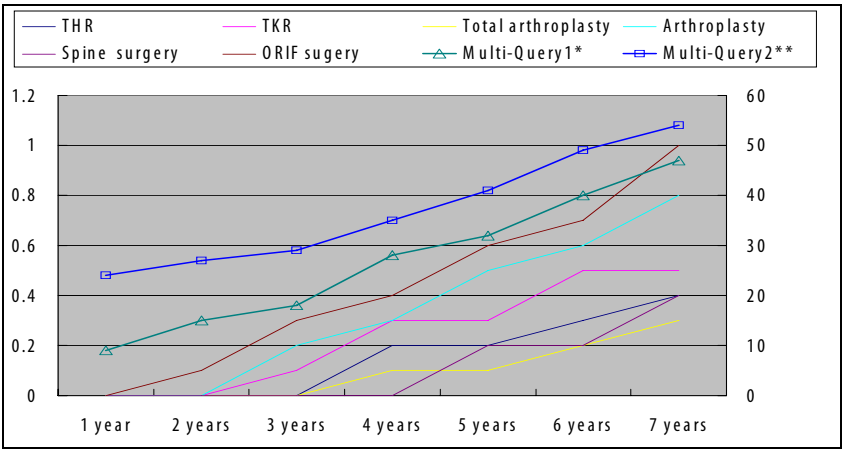


Fig. 2. Difference(in seconds) between local and remote executing efficiency. Multi-Query1 and Multi-Query 2 show their scale on the right side.

4 Conclusion and Disscusion

The system creates various and enormous products to meet users' requests. To achieve such purpose, knowledge architecture plays a crucial role. After the more brain-storming occurs, the more abundant knowledge architecture develops. If we expand the application range to other departments, that will develop a huge knowledge architecture and possess versatile functions.

The storage of medical information increases in an astonishing rate. Most of specialty database will face the obstacle of executing efficiency. The system use the

off-peak time to generate time-consuming tables and reports. That results some compromise of timeliness but make balance between efficiency and accuracy.

HIS is a huge machine and lack of reflex, therefore, invalid data accumulation is seldom detected by information administrators. Such invalid data may not result in financial loss but great medical statistics errors. Specific feedback from professional staffs is the best method to eliminate such errors. However medical staffs are not familiar with queries or data structures, they can not find the way to give feedback or explore the errors. The system does detect significant invalid data, which is eliminated for data integrity and make feedback to information administrators.

Incorrect medical behavior may cause serious consequence. The system can detect such behaviors and present the change if some improvement protocol has been executed. Continue quality control can be achieved through this process.

The productions of the system are abundant and variable. Some physicians of other department may question about the exact effect on decision-support. The impact on the decision-support can be evaluated through questionnaire evaluation. However the questionnaire can not comprise all aspects of decision-support. What we can improve is expand aspects of questionnaire evaluation. Besides, the system can form specific predict model [6] which is derived from answer of the specific question. The result of executing efficiency revealed a particular problem. The multiple queries executing time may approach one minute, which reflects necessity of SQL queries reformation and unsuitability of Access® for such application.

Future works -- As taxonomies of decision-support system, the system can be classified as a data-driven decision-support system. For architecture [7] of decision-support system, we lack a model component, which should be created through answering specific questions belonged to our knowledge architecture. That will spend time to accumulate the final answers to form rules and expand the aspects of application. The image is very important for orthopedic surgeons and incorporation of image database will enhance the effect of decision-support. The next step is upgrading the system through creation of image database and extraction image data from Picture Archiving and Communication System (PACs).

Conclusion -- This unique system for clinical and academic use of orthopedic departments, which possess functions of auto-extraction of patient data source from the hospital information system (HIS) and web interface for data queries. After a series of assessment and evaluation, the system do is able to play a positive role in the decision-support process.

References

1. McManus, J.G.: Orthopedic office medical records using COSTAR. Orthop. Clin. North Am. 17(4), 581–590 (1986)
2. Stoodley, M.A., Sikorski, J.M.: OASYS: a computerized auditing system for orthopaedic surgery. Int. J. Biomed. Comput. 29(2), 119–131 (1991)
3. Ruffin, M.: The importance of data warehouses for physician executives. Physician Exec. 20(11), 45–47 (1994)

4. Prather, J.C., Lobach, D.F., et al.: Medical data mining: knowledge discovery in a clinical data warehouse. In: Proc. AMIA Annu. Fall Symp., pp. 101–105 (1997)
5. Kerkri, E.M., Quantin, C., et al.: An approach for integrating heterogeneous information sources in a medical data warehouse. J. Med. Syst. 25(3), 167–176 (2001)
6. Lin, S.H.: Access the risk factors of transferring to intensive care units by artificial neural network in orthopedic practice. In: Annual meeting of orthopedic academy, Taiwan (2008)
7. Power, D.J.: Decision support systems: concepts and resources for managers. Westport, Conn., Quorum Books (2002)

Biometric-Gaussian-Stream (BGS) Cipher with New Aspect of Image Encryption (Data Hiding)

Maqsood Mahmud[1,2], Muhammad Khurram Khan[1],
and Khaled Alghathbar[1,2]

[1] Center of Excellence in Information Assurance, King Saud University, Riyadh
[2] Information Systems Department, College of Computer and Information Sciences,
King Saud University, Riyadh
maqsood.m@ksu.edu.sa, mkhurram@ksu.edu.sa,
kalghathbar@ksu.edu.sa

Abstract. Generally, a stream cipher is a symmetric key cipher where plaintext bits are combined with a pseudorandom Keystream bits to achieve desired cipher text. In our paper, we are proposing a cryptosystem named Biometric-Gaussian-Stream(BGS) cryptosystem. It is basically the combination of Gaussian noise and stream ciphers. In this system first, we pass an image through Gaussian noise function to add complexity to our system. This function is applied with specific parameters of mean and variance which works as a parallel key. To implement stream cipher with help of Biometric images, we extract Initial condition for LFSR from iriscode. We extracted iriscode using CASIA Iris database with help of Gabor Wavelet equation. A NASA image of "Saturn", with entropy of 3.97, was chosen for encryption and decryption purpose.

1 Introduction

Biometrics refers to methods for uniquely recognizing humans based upon one or more intrinsic physical or behavioral traits. In information technology, in particular, biometrics is used as a form of identity access management and access control. It is also used to identify individuals in groups that are under surveillance.[1],[2].

In cryptography, a stream cipher is a symmetric key cipher where plaintext bits are combined with a pseudorandom key bit stream (key stream), typically by an exclusive-or (xor) operation. In a stream cipher the plaintext digits are encrypted one at a time, and the transformation of successive digits varies during the encryption. An alternative name is a state cipher, as the encryption of each digit is dependent on the current state. In practice, the digits are typically single bits or bytes.[3],[4].

Stream ciphers represent a different approach to symmetric encryption from block ciphers. Block ciphers operate on large blocks of digits with a fixed, unvarying transformation. This distinction is not always clear-cut: in some modes of operation, a block cipher primitive is used in such a way that it acts effectively as a stream cipher. Stream ciphers typically execute at a higher speed than block ciphers and have lower hardware complexity. However, stream ciphers can be susceptible to serious security problems if used incorrectly. The same starting state must never be used twice.[3].

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 97–107, 2009.

Stream ciphers, where plaintext bits are combined with a cipher bit stream by an exclusive-or operation (xor), can be very secure if used properly. However they are vulnerable to attack if certain precautions are not followed:[5]

- keys must never be used twice
- valid encryption should never be relied on to indicate authenticity

In cryptography, the avalanche effect refers to a desirable property of cryptographic algorithms, typically block ciphers and cryptographic hash functions. The avalanche effect is evident if, when an input is changed slightly (for example, flipping a single bit) the output changes significantly (e.g., half the output bits flip). In the case of quality block ciphers, such a small change in either the key or the plaintext should cause a drastic change in the ciphertext. The actual term was first used by Horst Feistel (Feistel 1973), although the concept dates back to at least Shannon's diffusion [6], [25]. Diffusion dissipates statistical structure of plaintext over bulk of ciphertext. Confusion makes relationship between ciphertext and key as complex as possible.

2 Related Work

Paper stated in reference, [7],[9],[10], [12],[14]& [15] were found most relevant to our work. We have extend their work in the field of Biometrics. The other references [3]&[13]are based on stream ciphers. The ideas from both sets of references were combined to conceive the idea of Bio-Gaussian-Stream Cipher. In [10], the authors use user-specific biometric information instead of using PINS & passwords. They also present the generation of stable cryptographic keys from biometric data that are stable in nature. A longer and more stable biostream is generated as the Cryptogrphic key. The [14] proposes a novel two factor authentication based on iterated inner products between tokenized pseudo-random number and the user specific fingerprint feature, which is generated from the integrated wavelet and Fourier-Mellin transform, and hence produce a set of user specific component code that coined as Biohashing. In [7] the author describes the relationship between chaotic characteristic and cryptography, putting forward a chaotic algorithm of image encryption with double keys. Discrete logistics maps were used and were implemented in MATLAB like our cipher. The original value of the sequence was regarded as secret key. In [12] the authors talk about the recent researches on chaotic system. They further states that drawback of small key space and weak security in one-dimensional chaotic encryptions are obvious. In [8] the authors state about the biometric encryption. They think that password management is the weakest point of any cryptosystem, as the password can be guessed, found with a brute force search or stolen by an attacker. Moreover, because of variability the biometric image or template itself cannot serve as a cryptographic key. In [16] the authors present a novel chaos based cryptosystem to solve the privacy and security issues of biometric templates in remote biometric authentication over the network. Secret keys are randomly and dynamically generated without any human intervention, and each transaction session has different secret keys. Chaotic encryption scrambles the biometric templates into intangible form and chaotic modulation spreads the encrypted templates across a wide band of frequency. This makes them more difficult to decipher under attacks. The [9] describes a secure

fingerprint verification system based on the fuzzy vault scheme, where the sensitive biometric template is not stored, but a transformed version. It gives high unlock complexity for attackers with an acceptable unlock rate for the legal users is achieved. In [7] Chaotic algorithm of image encryption with double keys is described. It uses logistic maps to produce chaotic sequence, the original value of sequence is regenerated as secret key.

In this paper the Biometric-Gaussian-Stream cryptosystem gives a new dimension to the field of cryptography. This idea has been emerged by manipulating the existing ciphers and biometric security features and their strength in the current scenarios of insecure world of communication. The idea is the combination of three security aspects i.e. Biometrics (Iris),Gaussian Noise, Stream Ciphers. We used Gabor wavelet to generate iriscode for iris bits (Θ) generation. The use of smart card is to avoid the problem of variability in biometrics in our given case study. The smart card is considered to be already secured by any encryption techniques to avoid security breaches. The key used in the Bio-G-Stream Cipher is generated with help of biometrics (Iris code)[1],[18],[16],[17].This key is further used to encrypt multimedia objects like picture or even Iris image by self as is performed in our case. The selection of bits for the inputs LFSR[13] key is done by hamming code method. We used J.G. Daugman [1] proposed algorithm for the Iriscode generation. He used Gabor Wavelet equation to extract phase of the iris image. We further used these iris bits to extract specific bits using hamming method to feed LFSR[17],[20]. We added Gaussian noise using Gaussian function with variance (v) and mean LFSR [17].[20]. This Gaussian noise already added image is further passed through LFSR to encrypt it. The decryption was performed in the reverse order as usually done in cryptosystems. We take mean(m) and variance (v) as parallel keys in addition with Initial condition of LFSR.

3 Our Research Methodology

The following method was adapted to carry out our research.

i) Literature review on biometrics and existing stream ciphers
ii) Selection of any one Biometric feature like (finger print, palm geometry, speech, gait, iris etc).
iii) Finding Iriscode from Iris image. It is extracted using Gabor Wavelet equation and is given in Equation 1 [27].

$$G(x,y)=\frac{1}{2\pi\sigma\beta}\, e^{-\pi\left[\frac{(x-x_o)}{\sigma^2}+\frac{(y-y_o)}{\beta^2}\right]}\, e^{i\left[\xi_o x+v_o y\right]} \qquad (1)$$

where (*xo*, *yo*) is the center of the receptive field in the spatial domain and (ξ_o, v_o) is the optimal spatial frequency of the filter in the frequency domain. α and β are the standard deviations of the elliptical Gaussian along *x* and *y*. The 2D Gabor function is thus a product of an elliptical Gaussian and a complex plane wave.

iv) Extraction of bits (Initial Condition) to feed LFSR to generate key stream. We will use hamming method to extract bits positions. One can use any method like "Diagonal selection of bits", Double Cross Diagonal Selection of bits", "Zig Zag selection" or "Hash Table" etc. We are choosing bits position in iris template as follows:

$$\text{i.e. } 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, 2^8 \ldots\ldots\ldots$$

v) Implementing Stream cipher mechanism to encrypt the same or different image [19]. Passing the encrypted image through Gaussian Filter to make it too complex to comprehend. The Gaussian Noise used in this paper can be written as the probability density function of n-dimensional Gaussian noise [11],[26]:

$$f(x) = \left((2\pi)^n \det k\right)^{-1/2} \exp\left((x-\mu)^T k^{-1} \frac{(x-\mu)}{2}\right) \quad (2)$$

where x is a length-n vector, K is the n-by-n covariance matrix, μ is the mean value vector, and the superscript T indicates matrix transpose

vi) Finding Confusion and Diffusion elements
vii) At last to determine the avalanche Effect between (Plain text and Cipher Text), (Cipher text and Key) and (Plain text and Key)
viii) Comparison between RC2,RC4,DES,3DES and our Bio-G-Stream Cipher with respect to speed (Mbps) is also performed.

4 Proposed Biometric-Gaussian-Stream (BGS) Cipher Model

The Fig. 1 describes our proposed model for the Biometric-G-Stream (BGS) Cryptosystem. An iris image is taken and iris bits are generated using Gabor Wavelet equation. If the stored image in smart card[21] matches with live eye iris pattern then encryption will be done otherwise termination will be occurred (Policy based encryption to avoid the problem of variability).Note that smart card must already be secured with any strong algorithm (we are not concerned with security of smart cards).Then the stored iris template will be used for encryption. Initial Condition IC will be chosen using hamming method for LFSR to generate Keystream for encryption. The same or another image may be taken for encryption purpose as an object. Image bits are generated by imread() function in MATLAB. It is passed through Gaussian noise to make it more complex. Xor operation is applied between Keystream bits and image(Gaussian Noise added) bits. An encrypted image (Multimedia Hiding) is achieved. Now it could be securely transferred on channel. For decryption process the encrypted image is first XORed using Keystream from LFSR. The decrypted image is then passed into Gaussian function to have original image. These image bits can be reconstructed to original image by imshow(I) function in MATLAB. Note: the two dimensional aspects of the image is converted into one dimension and then transferred into binary format to get plain binary text to perform encryption.

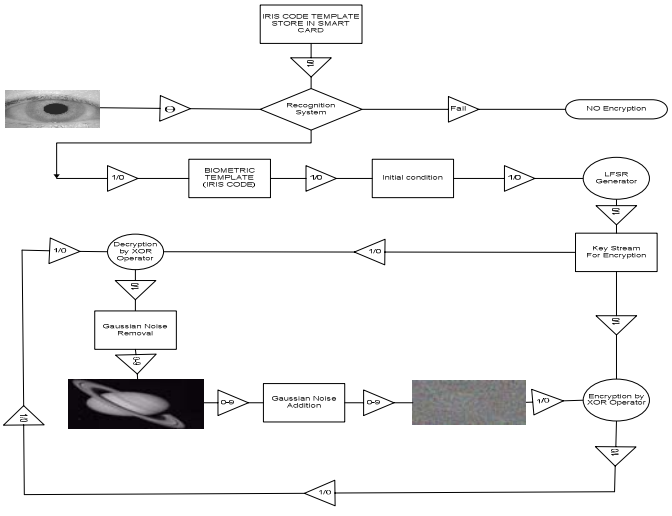


Fig. 1. Proposed Biometric-Gaussian-Stream (BGS) Cipher Model

5 Mechanism of Biometric-G-Stream Cipher (BGS)

The mechanism of BGS is described bellow with help of an example.

5.1 Description of Biometric-Gaussian-Stream Cipher Algorithm

The following example is given to give in depth understanding of the idea conceived regarding Bio-G-Stream ciphers. The algorithm is mentioned in section 6.1. An Iris code template is taken in binary. This image is converted into iriscode for the purpose of simulation in MATLAB [11]. We used Gabor Wavelet equation for iriscode generation. It is stored in biometric string named "b" in line 2. First initial condition is generated by hamming method (i.e. 2^0 , 2^1 ,2^2 ,2^3 ,2^4,2^5,2^6….) from the string "b" and stored in another string named as "IC" in line 4.

The LFSR[3] For loop is run to generate key stream for our bio-stream cipher. The loop starts in step 7 and end in step 10 to generate key stream.

In section 6.2 the Avalanche effect[5] (Shannon's Diffusion) is calculated using MATLAB[11] tool. A string "DiffPC" is declared in line 6.This strings will show the difference of bits in Plain Text and Cipher Text. In line 4 the DiffPC is calculated. A variable "Accumulator" is used for the summation of the difference in bits using Xor function. Percentage difference in Plain Text and Cipher text is calculated and stored in accumulator in line 11. Avalanche Effect between Cipher text and Key is also calculated and store in variable "AvalancheCK" and is shown in line 12. The Avalanche effect between Plain text and Key is shown in line 13. The value is stored

in variable “AvalanchePK”. In line 14, Avalanche effect is calculated and stored in variable “AvalanchePC”.

6 Implementation

The implementation of the Biometric-G-Stream is performed using MATLAB tool. The results are discussed in section 7.

6.1 Algorithm for Bio-Gaussian-Stream Cipher

The algorithm for stream cipher is mentioned below to have the in depth understanding of the idea. The code was written in MATLAB [11].

```
1   Begin
2   ► Input Biometric Data in
3   ► “Bio” variable
4   IC<-- Bio(1), Bio(2),Bio (4), Bio(8),Bio(16),Bio (32),Bio,(64),Bio (128)]
5   N<--128
6   ► LFSR Key Generation
7   For i<--1 to n
8   Key(i) <--IC(8)
9   IC(1)<--IC(2)XOR IC(5)11   Shift each bit forward
10  end
11  Read Image <-- Input Image
12  I <-- Image Data
13  Input <-- p
14  J<-- Gauss Noise(I, p)
15  CT<--I XOR Key
16  Decryption Process
17  PT<-- CT  XOR Key
18  Input<-- h
19  K <-- Gauss Filter(PT,h)
20  Show image <-- K
21  End
```

6.2 Avalanche Effect

The below algorithm is used for to evaluate the avalanche effect for the Biometric-Stream cipher. This code was written in MATLAB[11].

```
1   Begin
2   ► Finding the Avalanche
3   ► Effect/Shannon Diffusion
4   DiffCK<--Key XOR CT
```

```
5    DiffPK(i)<--PT XOR Key
6    DiffPC(i) <--PT XOR CT
7    Accum-->0
8    M<--128
9    For i<--1 to m
 Accum<-- DiffCK(i)+ Accum
 Accum<--DiffPK(i)+ Accum
 Accum<--DiffPC(i)+ Accum
10   end
11   PercCK<--Accum/128*100
12   AvalancheCK<--100-PercentCK
13   AvalanchePK<--100-PercPK
14   AvalanchePC<--100-PercPC
15   End
```

7 MATLAB Results/Findings

The results of our simulation in the perspective of Shannon's Diffusion or Avalanche's Effect can be viewed below.

7.1 Description of Table 1

The entropy is the randomness in a system. This is a special topic in the information theory and coding. We found the entropy of the system and found to be 64, which shows better randomness of bits with probability of 0.5 for each bits. More random is the cipher more strong it will be. Moreover the entropy of an image can be found easily using MATLAB command entropy(I); Generally Entropy is given by the Equation.2, [22].

$$H(X) = \sum_{i=1}^{n} p(x_i) \log_b p(x_i) \tag{3}$$

Where $p(x_i)$ is probbility of a single bit in a random stream of bits and is equal to 0.5.

Table 1. Entropy of Biometric Stream Cipher

Cipher	Entropy(H(X))
Biometric-Gaussian - Stream Cipher (BGS)	64

7.2 Description of Table 2

Table 2 describes the entropy of a specific image which is under our consideration. We took NASA picture to encrypt and found its entropy to be 3.9700. The entropy varies with picture, size and contents.

Table 2. Entropy of NASA Saturn Image

Image	Entropy (H(x))
NASA Image (Saturn)	3.9700

7.3 Description of Table 3

The Table 1 shows two different avalanche effects between two different variables i.e. (Plain Text with Cipher Text) and (Key Stream with Cipher text). The idea is conceived by the permutation phenomena where order matters like PK, PC, CK. First permutation takes two variable Plain text (P) and Cipher Text (C). The second permutation takes considers Plain Text (P) and Key Stream (K). The third permutation takes Cipher Text(C) and Key Stream (K). In table 3 we can see the AvalanchePC and AvalancheCK are near to 50 % change. This result can be further improved by using other biometric aspects or more complex stream cipher like Alternate step generating stream cipher or RC6 or A5 etc. The comparison is also shown in table 3 with DES. Since Avalanche effect of DES is bit higher then Biometric-Gaussian-Stream Cipher (BGS) but BGS is better in the sense that our cipher is light weighted with one XOR operation and Gaussian function which takes less time to encrypt and decrypt.

Table 3. Avalanche Effect (Shannon's Diffusion)

Cipher Name	AvalanchePC Effect for Bio-Stream Cipher	AvalancheCK Effect For Bio-Stream Cipher
Biometric-Gaussian-Stream Cipher	47.6563 %	50.7813 %
DES	53.125 %	54.6875 %

7.4 Description of Table 4

The table 4 depicts about the speed comparisons of Symmetric Ciphers. In BGS cipher the x* indicates the time taken by the Gaussian noise factor which is additional in comparison with RC4.We can conclude from the below table that our BGS is more robust then DES and 3DES with respect to speed.

Table 4. Speed Comparisons of Symmetric Ciphers

Cipher Name	Key Length	Speed (Mbps)
DES	56	9
3DES	168	3
RC2	variable	0.9
RC4	Variable	45
Biometric-Gaussian-Stream(BGS) Cipher	variable	45-x*

Where **x*** is a Gaussian Noise factor and it depicts decrease in speed(Mbps).

8 Limitation of Our System

Because of the variability, the biometric image or template itself cannot serve as a cryptographic key. We need to adopt a policy for encryption process in biometric images or templates. Our policy states that stored image template (Smart Card) will be useless for cryptanalyst until live matching (recognition) of human iris is matched. This means that specific threshold phenomenon for iris recognition will be only used for iris recognition, not for encryption while the stored template (Smart Card) will be used for encryption. Before matching or recognition the stored template (Smart Card) cannot be used for encryption purposed. This will solve partially or may be totally the problem of variability of biometric image or template itself.

9 Future Work

This work can be further extended to block cipher like DES & IDEA and advanced stream ciphers like eSTREAM. Chaotic Functions, Logistic Maps or Elliptic Curves can also brought into consideration with combination of biometric features to achieve more desired results not in the field of cryptography but also new emerging field of biometrics.

10 Conclusion

We developed a Biometric-Gaussian-Stream Cryptosystem with in this paper. Gaussian noise was added to the image with specific parameters to make it more complex. We extracted key from iriscode and encrypted the image with its help. The decryption was also made in the same fashion. This ciphers proves to be more strengthened in comparison with others ciphers like RC6 etc due to the biometric aspects. The Shannon's Diffusion or Avalanche effects shows Bio-Gaussian-Stream Cipher strength. The selection of iris code is due to its versatile behavior and its and uniqueness. The entropies of the cipher and image ware also found out and proves its strength. In conclusion, proposed system can be easily realized in the real environment and can easily replace A5/1 and RC6.

References

1. Irisbits pattern proposed by Dr. Daugman, http://www.cl.cam.ac.uk/~jgd1000/irisbits.txt
2. Journal of the International Biometric Society, http://www.biometrics.tibs.org/
3. Mahmud, M.: Natural Language (ARABIC) as a Strengthening Layer for Stream Ciphers in Wireless Networks. In: The 17th IASTED International Conference on Applied Simulation and Modeling, ASM 2008,Corfu Greece, June 23-25 (2008), http://www.actapress.com/PaperInfo.aspx?PaperID=33412&reason=500
4. Biham, E., Granboulan, L., Nguy, P.Q.: Impossible Fault Analysis of RC4 and Differential Fault Analysis of RC4. Computer Science Department, Technion – Israel Institute of Technology, Haifa
5. Avalanche Effect, http://en.wikipedia.org/wiki/Avalanche_effect (Date Accessed: 20 March 2009)
6. Stream Cipher Attacks, http://en.wikipedia.org/wiki/Stream_cipher_attack (Date Accessed: 19 March 2009)
7. Honge, R., Zhenwei, S., Yuanzhi, W., Jian, Z.: A Chaotic Algorithm of Image Encryption Based on Dispersion Sampling. In: The eight International conference on Electronic Measurement and Instruments. IEEE, Los Alamitos (2007)
8. Biometric Encryption, Biometrics and Cryptography, http://www.3Dface.org
9. Yang, S., Verbauwhede, I.M.: Secure Fuzzy Vault Based Fingerprint Verification System. IEEE, Los Alamitos (2004)
10. Chang, Y.-J., Zhang, W., Chen, T.: Biometrics-Based Cryptogrphic Key Generation. IEEE, USA (2004)
11. The MathWorksTM Accelerating the pace of engineering and science, http://www.mathworks.com (Date accessed: 2 February 2009)
12. Gao, H., Zhang, Y., Liang, S., Li, D.: A new Chaotic Algorithm for image Encryption. Elsevier, Science Direct (August 2005)
13. What is an LFSR,Texas Instrument SCTA036A (December 1996), http://www.pld.com.cn/freeip/LFSR.pdf
14. Jin, A.T.B., Ling, D.N.C., Goh, A.: Biohashing: two factor authentication featuring fingerprint data and tokenized random number. The Journal Of The Pattern Recognition Society (April 2004)
15. Paul, S., Preneel, B.: A new weakness in the RC4 keystream generator and an approach to improve the security of the cipher. In: Roy, B., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 245–259. Springer, Heidelberg (2004)
16. Khan, M.K., Zhang, J.: Implementing Templates Security in Remote Biometric Authentication Systems. In: IEEE Conf. Proceedings on CIS 2006, China, vol. 2, pp. 1396–1400 (2006)
17. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15, 1148–1161 (1993)
18. CASIA Iris Database (March 2006), http://sinobiometrics.com
19. Khan, M.K., Zhang, J.: Improving the Security of A Flexible Biometrics Remote User Authentication Scheme. Computer Standards and Interfaces (CSI) 29(1), 84–87 (2007)
20. Daugman, J.G.: Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters. J. Optical Soc. Amer. 2(7), 1160–1169 (1985)

21. Wang, X., Zhang, W., Zhang, J., Khan, M.K.: Cryptanalysis and Improvement on Two Efficient Remote User Authentication Schemes Using Smart Cards. Computer Standards and Interfaces (CSI) 29(5), 507–512 (2007)
22. Shannon, C.E.: A Mathematical Theory of Communication 27, 379–423, 623–656 (July-October 1948)
23. Horan, D., Guinee, R.: A Novel Stream Cipher for Cryptographic Applications. In: Military Communications Conference, MILCOM 2006, October 23-25, pp. 1–5 (2006)
24. Zhang, Y.-P., Sun, J., Zhang, X.: A stream cipher algorithm based on conventional encryption techniques. In: Canadian Conference on Electrical and Computer Engineering, vol. 2, pp. 649–652 (2004)
25. Schneier, B.: The uses and abuses of biometrics. Communications of the ACM 42, 136
26. Chapeau-Blondeau, F.: Nonlinear test statistic to improve signal detection in non-Gaussian noise. IEEE Signal Processing Letters 7, 205–207 (2000)
27. Lee, T.S.: Image Representation Using 2D Gabor Wavelets. IEEE Transactions On Pattern Analysis And Machine Intelligence 18(10) (October 1996)

DNA Microarray Classification Using Single Hidden-Layer Feedforward Networks Trained by SVD

Hieu Trung Huynh[1], Jung-Ja Kim[2], and Yonggwan Won[1]

[1] Department of Computer Engineering, Chonnam National university
300 Yongbong-dong, Buk-gu, Gwangju 500-757, Korea
hthieu@hcmut.edu.vn, ykwon@chonnam.ac.kr
[2] Division of Bionics and Bioinformatics, Chonbuk National University
664-14 St. #1 Dukjin-dong, Dukjin-gu, Chonbuk 561-756, Korea
jungjakim@chonbuk.ac.kr

Abstract. The development of computational algorithms in analyzing microarray data has attracted many researchers in recent years. Especially statistical and machine learning approaches can provide powerful tools for biomedical research such as gene expression interpretation, classification and prediction for cancer diagnosis, etc. In this paper, we investigate an application of SVD-Neural Classifier for microarray classification. The classifier is a single hidden-layer feedforward neural network (SLFN), of which the activation function of the hidden units is '*tansig*'. Its parameters are determined by Singular Value Decomposition (SVD). Experimental results show that the Neural-SVD model is simple, has low computational complexity and can produce better performance with compact network architecture.

Keywords: DNA microarray, classification, neural network, SVD, Singular Value Decomposition.

1 Introduction

High-throughput DNA microarray technology provides powerful tools in medical and biology fields. It offers the ability to study samples with small amount and to carry out the expression of thousands of genes at once. This makes possible to obtain vast amounts of information very quickly, from which the diagnosis and prognosis of disease based on the gene expression profiles can be well established. One of the typical applications of machine learning in computational biology is DNA microarray classification. The aim of this work is to identify patterns of expressed genes, from which it can predict class membership for new patterns.

There are many methods for classification tasks such as support vector machine (SVM), neural networks, or statistical techniques. However, the statistical techniques are improper because the microarray data is very high dimensional with the limited number of patterns and very little replication. SVM approach may take long time to select its model. The problem facing in applications of neural networks is training algorithms. Traditionally, training networks is based on gradient-descent algorithms

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 108–114, 2009.

which are generally slow and may get stuck in local minima. These problems have surmounted by extreme learning machine (ELM) algorithm which was recently proposed by Huang et al. [1] and is suitable for training single hidden-layer feedforward neural networks (SLFNs).

The salient point of ELM algorithm is that network parameters are determined by non-iterative steps, in which the input weights and hidden layer biases are randomly assigned and the output weights are determined by pseudo-inverse of hidden layer output matrix. This algorithm can obtain better performance with higher learning speed in many applications. However, it often requires a large number of hidden nodes which makes the trained networks respond slowly to new input patterns. These problems have been overcome by the improved ELM algorithms such as Least Squares Extreme Learning Machine (LS-ELM) [2], Regularized Least Squares Extreme Leaning Machine (RLS-ELM) [3], and Evolutionary Extreme Learning Machine (E-ELM) [4].

Recently, we introduced an effective training algorithm for SLFNs, of which the hidden layer activation function is '*tansig*' [5]. The classifier based on this training algorithm is called as *SVD-Neural classifier*, of which the parameters are determined by non-iterative simple steps based on Singular Value Decomposition (SVD) [5]. In this paper, we investigate an application of our SVD-Neural classifier for DNA microarray classification. Experimental results show that the SVD-Neural classifier can obtain better performance with fast learning speed.

The rest of this paper is organized as follows. Section 2 describes the SVD-Neural classifier and its application for DNA microarray analysis. Experimental results and analysis on real datasets for cancer diagnosis are shown in Section 3. Finally, we make conclusion in Section 4.

2 SVD-Neural Classifier

Recently, we introduced the *SVD-Neural classifier*, of which the hidden layer activation function is '*tansig*' and the parameters are determined by non-iterative simple steps based on Singular Value Decomposition (SVD) [5]. The architecture of the SVD-Neural classifier, which is the SLFN, is shown in Figure 1. It consists of P nodes in input layer, N nodes in hidden layer and C nodes in output layer. Unlikely to other neural network, the activation of hidden node must be '*tansig*' function defined by

$$f(x) = \frac{1-e^{2x}}{1+e^{2x}}. \tag{1}$$

Let $\mathbf{w}_m=[w_{m1}, w_{m2},\ldots, w_{mP}]^T$ be the input weights connecting from the input layer to the m-th hidden node, and b_m be its bias. The hidden layer output vector corresponding to the input pattern $\mathbf{x}_j$ is given by

$$\mathbf{h}_j=[f(\mathbf{w}_1\cdot\mathbf{x}_j+b_1), f(\mathbf{w}_2\cdot\mathbf{x}_j+b_2),\ldots, f(\mathbf{w}_N\cdot\mathbf{x}_j+b_N)]^T, \tag{2}$$

and the i-th output is given by

$$o_{ji}=\mathbf{h}_j\cdot\boldsymbol{\alpha}_i, \tag{3}$$

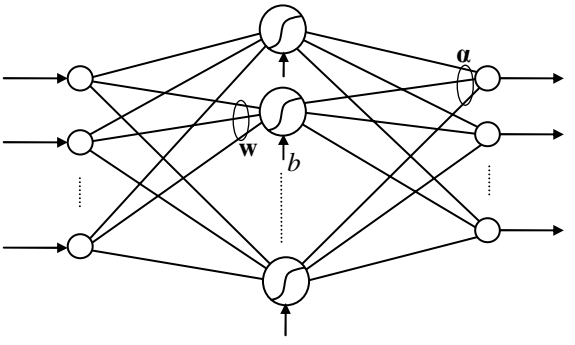


Fig. 1. Architecture of classifier

where $\boldsymbol{\alpha}_i=[\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iN}]^T$ is the weight vector connecting from the hidden nodes to the i-th output node.

Assume that there are n profiles or arrays in the training set $\mathbf{S}=(\mathbf{x}_j, \mathbf{t}_j)$, $j = 1, \ldots n$; each profile consists of P genes, $\mathbf{x}_j \in \mathbf{R}^P$; and let $\mathbf{t}_j=[t_{j1}, t_{j2},\ldots, t_{jC}]^T$ be the desired output corresponding to the input $\mathbf{x}_j$. We have to find parameters $\mathbf{G}$ consisting of input weights $\mathbf{w}$, biases b and output weights $\boldsymbol{\alpha}$ that minimize the error function defined by

$$E=\sum_{j=1}^{n}\left(\mathbf{o}_j-\mathbf{t}_j\right)^2=\sum_{j=1}^{n}\sum_{i=1}^{C}(\mathbf{h}_j\cdot\boldsymbol{\alpha}_i-t_{ji})^2 . \tag{4}$$

This is a least-squares approach of the linear problem defined by

$$\mathbf{HA}=\mathbf{T}, \tag{5}$$

where $\mathbf{H}$ is called the hidden layer output matrix and defined by:

$$\begin{aligned}\mathbf{H} &= [\mathbf{h}_1, \mathbf{h}_2,\ldots, \mathbf{h}_n]^T \\ &= \begin{bmatrix} f(\mathbf{w}_1\cdot\mathbf{x}_1+b_1) & \cdots & f(\mathbf{w}_N\cdot\mathbf{x}_1+b_N) \\ \vdots & \ddots & \vdots \\ f(\mathbf{w}_1\cdot\mathbf{x}_n+b_1) & \cdots & f(\mathbf{w}_N\cdot\mathbf{x}_n+b_N) \end{bmatrix},\end{aligned} \tag{6}$$

$$\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n]^T \tag{7}$$

and

$$\mathbf{A}= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2,\ldots, \boldsymbol{\alpha}_C]. \tag{8}$$

From (5), we can see that the matrix $\mathbf{T}$ is composed of the multiplication of two matrices $\mathbf{H}\in\mathbf{R}^{nxN}$ and $\mathbf{A}\in\mathbf{R}^{NxC}$. Thus, if we can reasonably decompose the matrix $\mathbf{T}$ into two matrices with sizes of nxN and NxC then parameters $\mathbf{G}$ of classifier can be determined simply.

In singular value decomposition (SVD) of $\mathbf{T}$, there exist an unitary matrix $\mathbf{U}\in\mathbf{R}^{nxn}$, a diagonal matrix with nonnegative real numbers $\mathbf{D}\in\mathbf{R}^{nxC}$, and an unitary matrix $\mathbf{V}\in\mathbf{R}^{CxC}$ so that

$$\mathbf{T}=\mathbf{UDV}^T \tag{9}$$

Let $\mathbf{U}=[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N, \ldots, \mathbf{u}_n]$, $\mathbf{D}=[\sigma_1, \sigma_2, \ldots, \sigma_n]^T$, and $\mathbf{V}=[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_C]^T$. Based on the properties of SVD decomposition, matrix $\mathbf{T}$ can be approximated by:

$$\mathbf{T} \approx \mathbf{U}_N \mathbf{D}_N \mathbf{V}^{\mathrm{T}}, \tag{10}$$

where $\mathbf{U}_N = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]$ and $\mathbf{D}_N = [\sigma_1, \sigma_2, \ldots, \sigma_N]^T$. Note that $\mathbf{U}_N \in \mathbf{R}^{nxN}$ and $\mathbf{D}_N \mathbf{V}^{\mathrm{T}} \in \mathbf{R}^{NxC}$. Therefore, matrices $\mathbf{H}$ and $\mathbf{A}$ can be determined by:

$$\mathbf{H}=\mathbf{U}_N \tag{11}$$

and

$$\mathbf{A}=\mathbf{D}_N \mathbf{V}^T. \tag{12}$$

Next, we have to determine the input weights and hidden layer biases. From (6) and (11), we have

$$\begin{bmatrix} f(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & f(\mathbf{w}_N \cdot \mathbf{x}_1 + b_N) \\ \vdots & \ddots & \vdots \\ f(\mathbf{w}_1 \cdot \mathbf{x}_n + b_1) & \cdots & f(\mathbf{w}_N \cdot \mathbf{x}_n + b_N) \end{bmatrix} = \mathbf{U}_N. \tag{13}$$

The input weights and hidden layer biases can be determined by using the linear system:

$$\mathbf{XW}=f^{-1}[\mathbf{U}_N]. \tag{14}$$

Note that $f^{-1}[\mathbf{U}_N]_{ij}=f^{-1}([\mathbf{U}_N]_{ij})$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}^T$$

and

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_N \\ b_1 & b_2 & \cdots & b_N \end{bmatrix}.$$

Finally, the minimum norm solution for $\mathbf{W}$ among all possible solutions is given by:

$$\hat{\mathbf{W}} = \mathbf{X}^{\dagger} f^{-1}[\mathbf{U}_N]. \tag{15}$$

In summary, the training algorithm for Neural-SVD classifier can be described as follows:

Given a training set $\boldsymbol{S}=\{(\mathbf{x}_j,\mathbf{t}_j) \mid j=1,\ldots,n\}$*, and number of hidden nodes N.*

1. Determine SVD of $\mathbf{T}$, and then calculate $\mathbf{U}_N$ and $\mathbf{D}_N$.
2. Determine the input weights and hidden layer biases by using (15).
3. Determine the output weights by using (12).

Thus, parameters of classifier can be simply determined by three steps. It is simple and has low computational complexity. It can obtain a compact network with small

number of hidden nodes and produce better performance for microarray data classification as shown in the following section.

3 Experimental Results

In this section, the performance of the SVD-Neural classifier is reported. The datasets for this study is described in Table 1. They have been widely used for the benchmark problems. They consist of two binary cancer classification problems: Leukemia data set [6] and colon cancer dataset [7]. The initial leukemia data set consisted of 38 bone marrow samples obtained from adult acute leukemia patients at the time of diagnosis, of which 11 suffer from acute myeloid leukemia (AML) and 27 suffer from acute lymphoblastic leukemia (ALL). An independent collection of 34 leukemia samples contained a broader range of samples: the specimens consisted of 24 bone marrow samples and 10 periferal blood samples were derived from both adults and children. The number of input features was 7,129. The objective is to separate the AML samples from the ALL samples. The training set consisted of 38 patterns and 34 patterns were used for testing.

The colon cancer data set contains the expression of the 2000 genes with highest minimal intensity across the 62 tissues derived from 40 tumor and 22 normal colon tissue samples [7]. The gene expression was analyzed with an Affymetrix (Sata Clara, CA U.S.A.) oligonucleotide array complementary to more than 6,500 human genes. The gene intensity has been derived from about 20 feature pairs that correspond to the gene on the DNA microarray chip by using a filtering process. Details for data collection methods and procedures are described in [7], and the data set is available from the website *http://microarray.princeton.edu/oncology/.*

The average results of fifty trials on two data sets are shown in Table 2 and Table 3, which shows comparison results of SVD-neural classifier approach with other training algorithms for SLFN such as ELM, RLS-ELM, and back-propagation (BP). The training algorithms were implemented in MATLAB 7.0 environment. The input features were normalized into the range [-1, 1]. The number of hidden nodes was gradually increased by 1 to find the near-optimal number of nodes based on cross-validation method.

From Table 2, we can see that the number of hidden nodes for SVD-Neural classifier was 2 for two data sets, which is equal to that of RLS-ELM and BP algorithm, and is about 10 times smaller than that of ELM algorithm. As shown in Table 3, the SVD-Neural classifier can obtain the classification accuracies of 95.93% and 83.63% for testing sets of Leukimia and Colon datasets, respectively. This show that the SVD-Neural classifier is comparable to RLS-ELM while outperforming classifiers based ELM, and BP algorithms.

Table 1. Specification of the data sets

Data set	Training set	Testing set	Gene expression levels
Leukemia	38	34	7,129
Colon	40	22	2,000

Table 2. Architecture comparison of SLFN training algorithms

Dataset	Methods	# hidden nodes
Leukimia	SVD-Neural	2
	RLS-ELM	2
	ELM	20
	BP	2
Colon	SVD-Neural	2
	RLS-ELM	2
	ELM	20
	BP	2

Table 3. Performance comparison with other classification methods (%)

Dataset	Method	Training		Testing	
		Mean	Dev.	Mean	Dev.
Leukimia	SVD-Neural	**100**	**0.00**	**95.93**	**5.11**
	RLS-ELM [3]	100	0.00	95.60	4.45
	ELM [3]	91.35	4.78	67.65	11.93
	BP [3]	98.85	9.96	88.52	14.36
Colon	SVD-Neural	**100**	**0.00**	**83.63**	**6.15**
	RLS-ELM [8]	99.75	0.79	83.27	6.61
	ELM [8]	88.35	5.06	64.18	10.50
	BP	95.70	10.45	80.27	10.53

4 Conclusion

Microarray classification is one of the typical applications in computational biology using the machine learning approaches. It can help to identify patterns of expressed genes from which class membership for new patterns can be predicted. In this paper, we introduce SVD-Neural classifier and investigate the performance comparison with other popular training algorithms for SLFNs. In SVD-Neural classifier, the SLFN has '*tansig*' activation function and the parameters of the classifier are determined by Singular Value Decomposition (SVD) approach. This training approach is simple and requires low computational complexity.

The performance of the training algorithms for SLFN such as SV- approach, RLS-ELM, ELM, and BP was evaluated and compared in terms of the number of hidden nodes and classification accuracy on test data sets. Data sets we used for this study were two binary cancer data sets of DNA microarray: leukemia and colon cancers. SVD approach, RLS-ELM and BP algorithms require the same number of hidden nodes, while ELM needs more hidden nodes. For classification accuracy, SVD-approach and RLS-ELM algorithms are comparable each other, while better than ELM and BP.

References

1. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501 (2006)
2. Huynh, H.T., Won, Y.: Small number of hidden units for ELM with two-stage linear model. IEICE Trans. on Information and Systems E91-D(4), 1042–1049 (2008)
3. Huynh, H.T., Won, Y., Kim, J.-J.: An improvement of extreme learning machine for compact single-hidden-layer feedforward neural networks. International journal of neural systems 18(5), 433–441 (2008)
4. Zhu, Q.-Y., Qin, A.K., Suganthan, P.N., Huang, G.-B.: Evolutionary Extreme Learning Machine. Pattern Recognition 38, 1759–1763 (2005)
5. Huynh, H.T., Won, Y.: Training Single Hidden Layer Feedforward Neural Networks by Using Singular Value Decomposition. In: Submitted to ICCIT 2009 (2009)
6. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science 286(5439), 531–537 (1999)
7. Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J.: Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. of National Academy of Sci. of the USA 96, 6745–6750 (1999)
8. Huynh, H.T., Won, Y., Kim, J.-J.: DNA microarray classification with compact single hidden-layer feedforward neural networks. In: Proc. of the frontiers in the convergence of bioscience and information technologies (FBIT 2007), pp. 193–198 (2007)

The Relationship Analysis of Skin Physiology Factors via Grey Theory

Ya-Ting Lee[1] and Chian-Song Chiu[2,*]

[1] Department of Beauty Science, Chienkuo Technology University, Changhua 500, Taiwan
ytlee@ctu.edu.tw
[2] Department of Electrical Engineering, Chung-Yuan Christian University, Chung-Li 32023, Taiwan
cschiu@dec.ee.cycu.edu.tw

Abstract. This research focuses on the skin physiological factor relationship analysis using grey model, namely GM(1, N) and GM(0, N). First of all, skin physiological factor sampling are done for Taiwan female subjects aged from 18 ~ 52. Here, we applied the GM(1, N) and GM(0, N) model to obtain the relationship weighting between the major factor and the other relational factors. Furthermore, with the determined weightings, we proceed on the skin physiological factor relationship analysis to understand the skin characteristics under different age.

Keywords: Grey relational analysis, skin physiological factors.

1 Introduction

Detailed skin condition is undetectable under human eye, where accurate instruments are required for physiological detections and skin status detections [1]-[3]. Current clinical skin detections mostly focus on simple analysis, yet they lack of complete skin diagnose and skin physiological analysis under different age. This is because skin physiological analysis requires large amount of sample data. Then, traditional methods lead to great difficulties in clinical researches.

Grey system theory includes grey relational analysis (GRA), globalization grey relational grade (GGRG), grey model (GM) [4]-[8]. In most of the researches, grey theory takes an important part in theoretical analysis and numerical data modeling and analysis. Furthermore, the constructed data model is a non-functional type sequence model with characteristics of easy computing and the numerical data do not need to satisfy classical distribution.

Based on the skin physiological analysis and grey theory application researches, we here establish the GM(1, N) and GM(0, N) model for relationship analysis of the skin physiological factors. First, experimental data base is constructed according to the samples from different subjects including the following 6 factors: age, skin elasticity, skin pH value, skin pigmentation, skin surface lipids and skin epidermal hydration. We arrive with the following conclusions:

* Correspondence addressee.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 115–122, 2009.

(1) Along with the increase of age, skin elasticity decreases rapidly, pH value decreases (skin tends to acid), and skin epidermal hydration increases causing the skin to lose moisture capacity.
(2) To maintain skin elasticity, it is necessary to retain the skin pH value and epidermal hydration. Besides, the more moisture is kept inside the skin, the better the skin elasticity.

2 Research Methods

2.1 Skin Physiological Detection

(1) Detection condition: Ambient temperature at 20 ± 2℃, and relative humidity at 45 ±5%.

(2) Age and sex: 18 – 52 Taiwan females.

(3) Tested skin partition: With different skin partition, accompanies with different epidermal hydration, surface lipids and pigment, and etc. In this research, we measured on both sides of the cheeks 3 times for each subject.

(4) Cleaning of the tested skin partition: Before detection, we use non-alcoholic scrub for cleaning and waited 3 hours after cleaning for the skin to recover.

(5) Skin features: We test dry and oily skin half and half of the subjects 3 times individually. Skin sensitive subjects are not taken into consideration.

In this research, 61 Taiwan female subjects are tested where the results are presented in Table 1.

Table 1. Original tested data

Age	Elasticity	pH	Surface Lipids	Pigmentation	Epidermal Hydration
20	0.8	4.5	278	200	15
23	0.86	5.7	143	167	14.5
25	0.82	4.82	136	279	13
…	...	…	…	…	…
29	0.78	4.8	178	178	20
31	0.77	4.75	148	230	17
33	0.88	5.33	185	156	15
39	0.87	4.88	82	199	24
42	0.77	5.2	110	380	21

2.2 Grey Theory

A. GM(1, N) Model

First we consider that the data can be separated into two sequences: (1) Major sequence factor, which is the sequence that masters the system behaviors; (2)

Influencing sequence factors, which are the sequences that influence the system behaviors. Here, the major sequence factor is defined as follows:

$$x_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \cdots, x_1^{(0)}(M))$$

where M is the sequence length. The influence sequence factors are

$$\begin{aligned} x_2^{(0)} &= (x_2^{(0)}(1), x_2^{(0)}(2), \cdots, x_2^{(0)}(M)) \\ x_3^{(0)} &= (x_3^{(0)}(1), x_3^{(0)}(2), \cdots, x_3^{(0)}(M)) \\ &\vdots \\ x_N^{(0)} &= (x_N^{(0)}(1), x_N^{(0)}(2), \cdots, x_N^{(0)}(M)) \end{aligned}$$

In other words, $x_1^{(0)}$ and $x_i^{(0)}$ represents the major sequence and the influence sequence. Then, $x_i^{(1)}$ and $x_i^{(0)}$ are defined as 1-AGO sequences $(i = 1, 2 \cdots, N)$:

$$x_i^{(1)} = \left(\sum_{k=1}^{1} x_i^{(0)}(k), \sum_{k=1}^{2} x_i^{(0)}(k), \cdots, \sum_{k=1}^{M} x_i^{(0)}(k) \right)$$

Let $z_1^{(1)}$ is the average generation of adjacent data sequence of $x_1^{(1)}$:

$$z_1^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1) \quad , \; k \geq 2$$

The grey differential equations of the GM(1,N) model is：

$$\frac{d\,x^{(1)}(k)}{d\,t} + ax_1^{(1)}(k) = \sum_{i=2}^{N} b_i x_i^{(1)}(k)$$

where a and b_i are determined coefficients; k = 1, 2,…, M. According to the GM(1,N) model form, the constructed 1-AGO sequence is:

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{i=2}^{N} b_i x_i^{(1)}(k) \tag{1}$$

The above equation (1) is the GM(1, N) model. Thus, we can arrive

$$\begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(M) \end{bmatrix} = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \cdots & & \\ -z_1^{(1)}(M) & x_2^{(1)}(M) & \cdots & x_N^{(1)}(M) \end{bmatrix} \begin{bmatrix} a \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$

Thus, $\hat{a} = [a, b_2, \cdots b_N]^T$ can be obtained using least square method where the details are as follows.

Theorem 1. Assume $x_1^{(0)}$ is the system major sequence, $x_i^{(0)}$ $(i = 2\cdots, N)$ is the influencing sequences, $x_i^{(1)}$ is the 1-AGO sequence, $z_1^{(1)}$ and $x_1^{(1)}$ is the average generation of adjacent data sequence. Thus the least square estimation of $\hat{a} = [a, b_2, \cdots b_N]^T$ is

$$\hat{a} = (B^T B)^{-1} B^T Y$$

where

$$Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(M) \end{bmatrix}, B = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \cdots & \cdots & \cdots \\ -z_1^{(1)}(M) & x_2^{(1)}(M) & \cdots & x_N^{(1)}(M) \end{bmatrix}$$

■

B. GM(0, N) Model

GM (0, N) model is a special case of GM(1, N) model with no derivatives which is classified as static factor analysis. The GM(0, N) model is as follows.

$$\begin{aligned} x_1^{(1)}(k) &= a + \sum_{j=2}^{N} b_j x_j^{(1)}(k) \\ &= a + b_2 x_2^{(1)}(k) + b_3 x_3^{(1)}(k) + \cdots + b_N x_N^{(1)}(k) \end{aligned} \tag{2}$$

where $x_i^{(1)}$ is the 1-AGO sequence of every $x_i^{(0)}$ $(i = 1, 2\cdots, N)$. If we submit all the sequence factors into (2) where $k = 1, 2,\ldots, M$, we can arrive with

$$\begin{bmatrix} x_1^{(1)}(2) \\ x_1^{(1)}(3) \\ \vdots \\ x_1^{(1)}(M) \end{bmatrix} = \begin{bmatrix} 1 & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ 1 & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \cdots & & \\ 1 & x_2^{(1)}(M) & \cdots & x_N^{(1)}(M) \end{bmatrix} \begin{bmatrix} a \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$$

Hence, using least square method solves the parameter vector $\hat{a} = [a, b_2, \cdots b_N]^T$.

Theorem 2: Let $x_1^{(0)}$ be the major sequence, $x_i^{(0)}$ $(i = 2, 3, \cdots, N)$ be the influencing sequences, and $x_i^{(1)}$ is the 1-AGO sequence of $x_i^{(0)}$ for $i = 1, 2, \cdots, N$. Therefore, the least square estimation of $\hat{a} = [a, b_2, \cdots b_N]^T$ is

$$\hat{a} = (B_0^T B_0)^{-1} B_0^T Y_0$$

where

$$B_0 = \begin{bmatrix} x_2^{(1)}(2) & x_3^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ x_2^{(1)}(3) & x_3^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \cdots & \cdots & \cdots & \cdots \\ x_2^{(1)}(M) & x_3^{(1)}(M) & \cdots & x_N^{(1)}(M) \end{bmatrix}, \quad Y_0 = \begin{bmatrix} x_1^{(1)}(2) \\ x_1^{(1)}(3) \\ \vdots \\ x_1^{(1)}(M) \end{bmatrix} \quad \blacksquare$$

Therefore, grey model structure of the influence sequences and major sequence can be represented as Fig. 1.

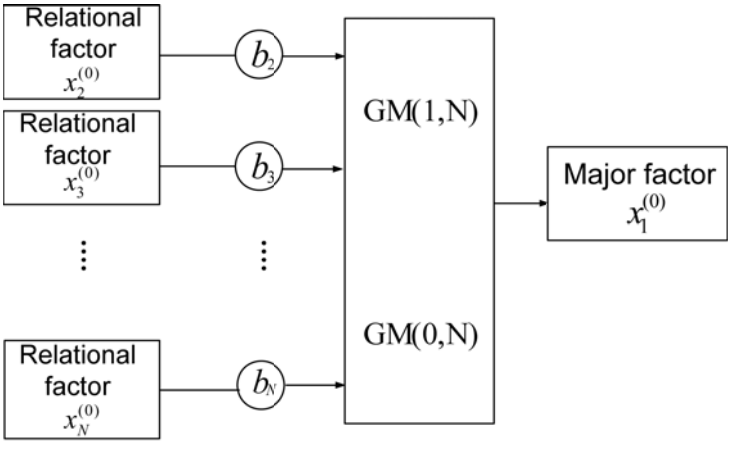


Fig. 1. Block diagram of the relationship analysis

C. Relationship Analysis

In the mentioned GM(1, N) and GM(0, N) model, grey system can be considered as an approximation model with input and output sequences, where the relationship between the two sequences are derived though the system parameter vector $\hat{a}$ for analyzing the relationship between age and the other physiological factors.

3 Analysis Results

First, we proceed on the age relational factor analysis where the structure is shown as Fig. 1. Here, age is defined as major sequence and submitted into GM(0, N) and GM(1, N) model to find out the relationship weighting b_j between the major sequence and the other sequences. The analysis results are given in Fig. 2 and we can see that age to skin elasticity, skin pH value, and skin epidermal hydration have the highest relativity where the weighting increases with its influence. From the weighting factors, it is obvious that the skin elasticity and pH value decrease with age. Moreover, the skin epidermal hydration increases with age causing the skin to lose its moisture capability. For more detailed results, the average data are given in Table 2.

	Elasticity	PH Value	Epidermal hydration	Surface Lipids	Pigmentation
GM(0,N)	30.8782	2.2461	0.8167	0.0569	0.0301
GM(1,N)	18.9117	5.0822	1.1289	0.0027	0.035

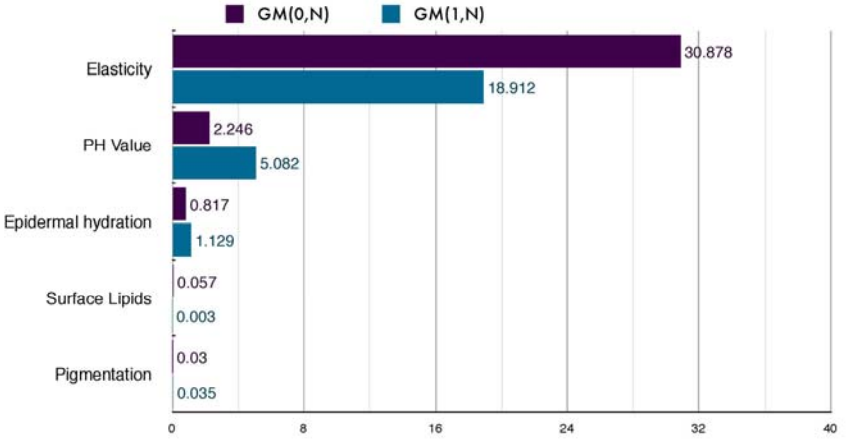


Fig. 2. Age as the major sequence

Table 2. Average value of individual factors

Average	Elasticity	pH	Surface Lipids	Pigmentation	Epidermal Hydration
Under age 25	0.81675	5.24775	172.225	230.25	16.8475
Above age 25	0.785714	5.098095	130.619	267.7143	21.1

From Table 2, the relationship between the skin elasticity (or pH value) and age is

$$\text{Age below } 25 > \text{Age over } 25$$

The relationship between the skin epidermal hydration and age:

$$\text{Age below } 25 < \text{Age over } 25$$

To carry on the research, we neglect the age influence and focus on the relationship between skin elasticity and pH value to the other skin physiology factors. We remove the age factor and separate the age groups into above and below 25 years old, then, let skin elasticity as the major sequence. The relational factors is obtained by using GM(1, N) and GM(0, N) model, where the results are given in Fig. 3 and Fig. 4.

According to the above results, skin elasticity, pH value, and epidermal hydration have greater relevance in the two age groups. Meaning that in order to maintain skin elasticity, it is necessary to maintain skin pH value (the pH value is direct proportion to skin elasticity) and reduce skin epidermal hydration. The less skin epidermal hydration, the more moisture is kept inside the skin and the better skin elasticity is obtained.

	PH Value	Epidermal hydration	Surface Lipids	Pigmentation
GM(0,N)	0.0775	0.0044	0.0024	0.0004
GM(1,N)	0.1101	0.0163	0.0033	0.0012

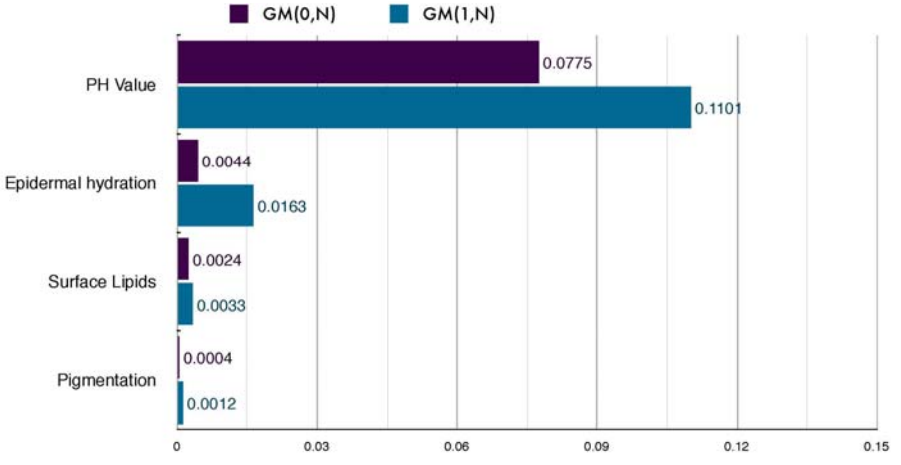


Fig. 3. Skin elasticity weighting for age under 25

	PH Value	Epidermal hydration	Surface Lipids	Pigmentation
GM(0,N)	0.1231	0.0021	0.0012	0.0002
GM(1,N)	0.1718	0.0298	0.0017	0.0018

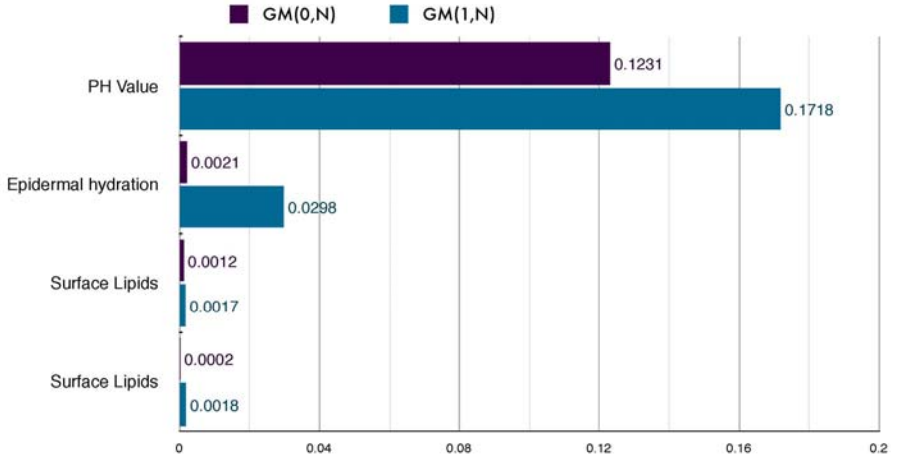


Fig. 4. Skin elasticity weighting for age above 25

4 Conclusions

Traditional skin physiology estimation method uses statistics analysis where large data samples are required. With restricted conditions, it can cause the clinical research results lack of authenticity and unsuitable to apply in practical use. By using grey system modeling, the relationship of the physiology factors can be indirectly obtained along with easy computing such that the disadvantages of traditional method are overcome. Moreover, the result is contributory for skin physiology relationship analysis and provides future decision-making in skin preservation.

Acknowledgments. This work was supported by the National Science Council, R.O.C., under Grant NSC-97-2221-E-033-059 and Ministry of Economic Affairs, DoIT, R.O.C., under Grant 98-EC-17-A-07-S2-0029.

References

1. Black, D., Del Pozo, A., Gall, Y.: Seasonal and anatomical variations in the surface state of the stratum corneum. Journal of Investigative Dermatology 108, 824 (1997)
2. Sator, P.G., Schmidt, J.B., Hönigsmann, H.: Comparison of epidermal hydration and skin surface lipids in healthy individuals and in patients with atopic dermatitis. Journal of the American Academy of Dermatology 48, 352–358 (2003)
3. Tupker, R.A., Schuur, J., Coenraads, P.J.: Irritancy of antiseptics tested by repeated open exposures on the human skin, evaluated by non-invasive methods. Occupational Health and Industrial Medicine 38, 83–84 (1998)
4. Deng, J.L.: Grey system theory and applications. Pearson, London (2003)
5. Wen, K.L.: Grey systems: modeling and prediction. Yang's Scientific Research Institute, USA (2004)
6. Tupker, R.A., Schuur, J., Coenraads, P.J.: Irritancy of antiseptics tested by repeated open exposures on the human skin, evaluated by non-invasive methods. Occupational Health and Industrial Medicine 38, 83–84 (2003)
7. Xue, H.B.: A further optimization in an optimized grey GM(1,1) model. Journal of Grey System 11, 107–112 (2008)
8. Zhou, P.W.: The optimization of background value in grey model GM(1,1). Journal of Grey System 9, 139–142 (2006)
9. Proceeding of The 2002 (7th) National Conference on Grey System Theory and Application

Advanced Analysis Information Architecture for Dosage Evaluation of Morphine Combining Adjunct

Hsiao-Hsien Rau, Wei-Tse Tang, and Chien-Yeh Hsu

Graduate Institute of Biomedical Informatiocs in Taipei Medical University,
No.250, Wuxing St., Xinyi Dist., Taipei City
cuhsu@tmu.edu.tw

Abstract. According to the statistics provided by National Bureau of Controlled Drugs in Taiwan and the United Nations International Narcotics Control Board, the DDD (Defined Daily Doses) of morphine consumption per million inhabitants in Taiwan was 12.2 mg in 1987, and substantially increased to 257.7 mg in 2002, which has been growing by 20 times in 16 years. The average daily consumption of morphine in Taiwan is one-twelfth of in United States. [1]In this study, we use the National Health Insurance Research Database (NHIRD) as our data source. The data was then pre-processed and saved into the database. We also developed an online inquiry system, which allows users to setup their own search terms and query outpatient data for follow-up analysis. In the future, we will provide more search methods by using other terms such as ICD-9 (International Statistical Classification of Diseases) for inquiries. In addition, the study will analyze not only the use of morphine and their combined drugs, but also establish Bayesian network for establishing the relationship between parameters, and analyze the effect of combined drug on morphine dosage.

1 Introduction

1.1 Morphine-Type Drugs

Morphine is a type of opium drug, which is a kind of alkaloid refined from poppy and provides strong analgesic effect. It gives significant effect on dread disease, trauma, surgery and also pain control for cancer patients. The initial use of morphine is by oral. Doctors in the United Kingdom in 1853 invented hypodermic needles, and in 1858, Doctors in the United States found the way to inject morphine directly into the vein. In 1950, the United States in order to artificially synthesized morphine, which is more effective, and less side effects [2].

Through binding with the receptor in gastrointestinal tract and central nervous system, neuron was hyperpolarized by opium, thereby inhibiting the stimulation of neural action potential and catalyzes the releasing of neurotransmitters from presynaptic inhibition. This made it participated in the regulation of spinal pain indirectly, and stimulates the releasing of glutamate from nerve endings.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 123–129, 2009.

In 1984, the World Health Organization (WHO) recommended oral morphine as the standard drug in the treatment of severe cancer pain, and developed the "By three Ladder" allay plan of cancer pain, including: ①. Non-opioid analgesics ± Supplementary analgesic drugs (anti-spasm, anti-anxiety, antidepressant and antipsychotic) ②. Weak opioid analgesics (codeine class drugs) + non-opioid analgesic ± analgesia ③. Powerful opioid analgesics (morphine class drugs) + non-opioid analgesic ± analgesia [3].

1.2 Morphine-Type Drugs and Cancer Pain Management

Currently, the treatment of cancer pain is mainly based on the guidelines formulated by WHO.[6] Started from the non-steroidal anti-inflammatory drugs (NSAIDs) with a weaker analgesic effect, along with the severity of pain increases, then slowly changed to a stronger opiate pain medication as the primary analgesic. Although opiate analgesic is currently a main treatment of cancer pain, but the reason for causing pain is very complicated. It is often unable to achieve adequate analgesic effect[7]. In addition, opiate analgesics also give rise to a number of intolerable side effects [8], such as skin itching, nausea, vomiting, constipation, lethargy and so on. This makes cancer patients unwilling to continue treatment, thus limiting its treatment effect. Therefore, combined use of opioids and appropriate analgesics, is a common way [9]. The analgesics which often used with opioids are supporting NSAIDs, tricyclic antidepressant, anxiolytics, anticonvulsants, oral local anesthetics, corticosteroids, antispasmodics and so on. These drugs can be used to treat different types of pain. And with the combined use of opioids, it can not only enhance the analgesic effect, but also reduce the use amount and side effects of opioids, and improve the symptoms caused by cancer. Although these drugs are widely used in cancer pain, but NSAIDs and other ancillary drugs will still causing considerable side effects, thereby affecting the effects and patient willingness of analgesics. Therefore, it is very helpful with cancer pain treatment if we can find analgesics and adjuvant which are more effectible but less side effects.

2 Methods

2.1 Database Framework

Currently, there are two large databases which are relevant with National Health Insurance (NHI). One is the original database system of NHI in Bureau of National Health Insurance, it storage all data related with diagnosis, medication and reports from NHI. In addition, since year 2000, Bureau of National Health Insurance commissioned National Health Research Institutes (NHRI) to issue the National Health Insurance Research Database (NHIRD), which contains all the records of NHI from start. Its data content is integrity, and researchers can apply for different type of information for research depends on their subject and purpose. The information our academic research departments retrieved so far are all provided by the database.

Our study collected 10% samples of the "Medical Claim Dataset" from 2000 to 2005, then divides it into outpatient and inpatient two parts, and saved in separate

database individually according to its year. Each database was then classified into "Diagnoses" and "Index", and saved the 366 tables which have been categorized by birthdates (Jan. 1st to Dec. 31st) into MySQL database on the server.

In order to have a more pure raw data, and simplify how the data presents, the study also carried out some pre-processing to the original data. We summarized all prescription records of each person into one record, and merged ORDER_TYPE, DRUG_NO, DRUG_USE, DRUG_FRE, UNIT_PRICE these six columns, each column is divided by ":". We then merged all the prescription contents into one column, and each record is divided by "|".

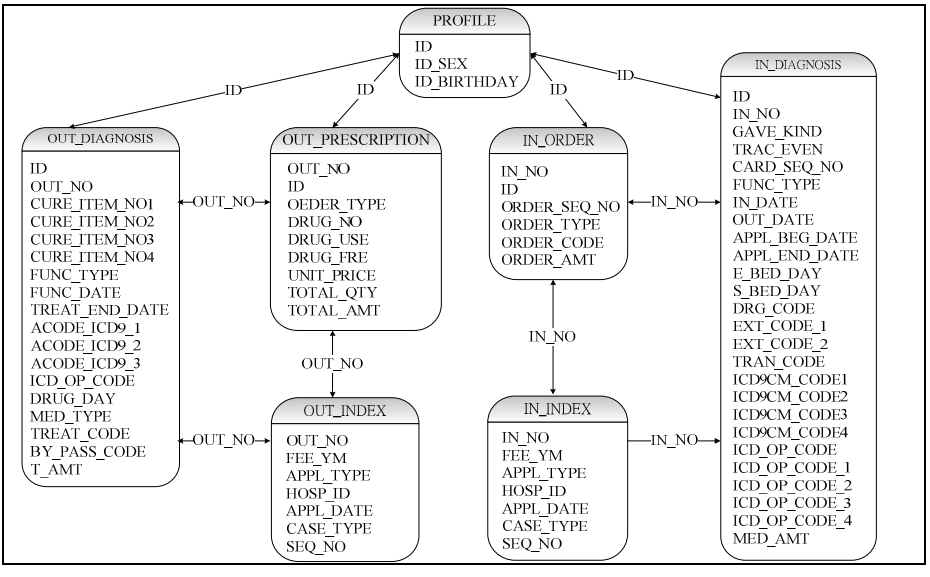


Fig. 1. Database Structure

This study considered the user needs have extended searching and server loading of queries, created a database named "Result" apart from the above database structure. Each result of a query will be saved as an independent table depends on the user, so the next query can use these tables as its data resource to reduce the amount of queries.

2.2 Application Development

We used Microsoft Visual Studio 2008 as our development tools in this study and VB.NET programming language for our web version NHIRD query tool. Considering the needs of users, our system is designed to be a flexible query platform based on demands. The functions in this study can adjust beginning and ending years, select or add drug codes in the dropdown menu, and set the query subject to inpatients or outpatients. After the query is done, we also provided function for user to export the result as Excel format. The user interface of the system is shown in Fig. 2.

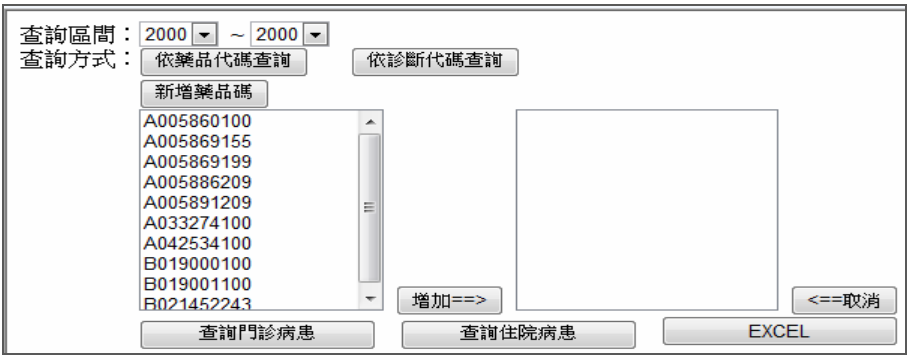


Fig. 2. The User Interface of System

2.3 Tool of Data Analyzing

We used the querying system to create data subset depending on different time periods and drugs. Descriptive statistics and Bayesian network model was constructed by using WEKA 3.6.0.

WEKA is a data mining tool using JAVA as its development environment, and was first proposed by the research team of University of Waikato in New Zealand. Including visualized data collecting tools and a lot of data analysis models and algorithms, WEKA supports multiple standards of data mining tasks, which includes data pre-processing, clustering, classification and regression etc. In addition, WEKA also provides JAVA database connectivity for user to link and query SQL database, and set the query results as input sources.

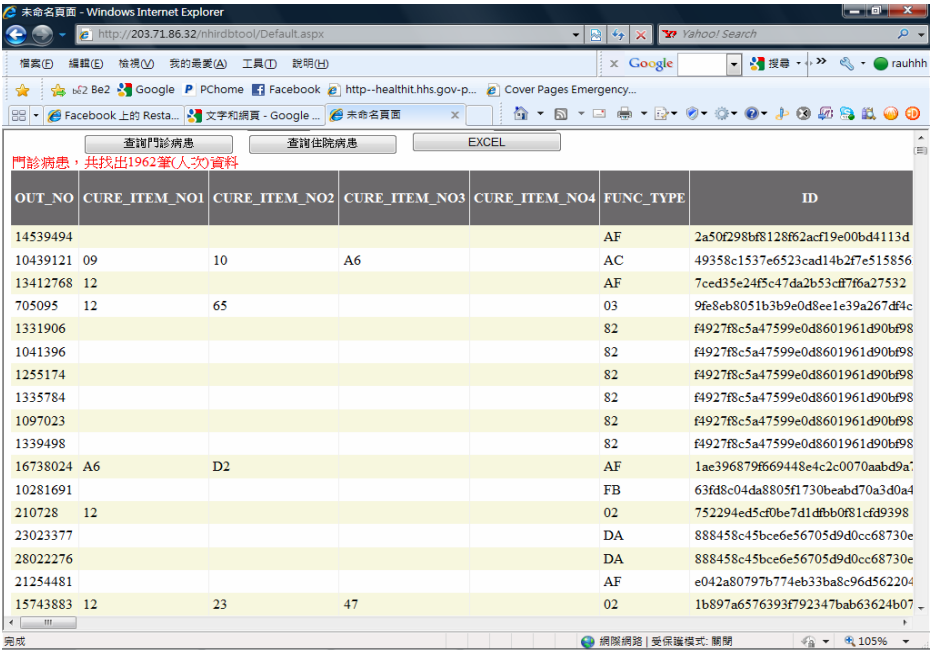
3 Results

3.1 Database Application

The main purpose of this system is to enable users to easily query all the medication record which meet the search limits form NHIRD. Fig. 3 shows the screen display of query results of the medical records that involved using morphine from 6-year NHIRD.. In this query, the user specified the beginning year, ending year, drug codes or ICD-9 codes. The application so far provides inpatients and outpatient records for query, and presents the results by using ASP.NET Data Grid when the query is done. It can also save every former result as a new table for better performance and shorten the query time if the same query results are to be used more than once.

About query performance, our study has run several tests and evaluations with different query intervals and a number of drugs. The results are as follows.

The above table shows that if the query interval reached three years or more than three years, it requires a longer time to finish, about 15 – 20 minutes in average. After



OUT_NO	CURE_ITEM_NO1	CURE_ITEM_NO2	CURE_ITEM_NO3	CURE_ITEM_NO4	FUNC_TYPE	ID
14539494					AF	2a50f298bf8128f62acf19e00bd4113d
10439121	09	10	A6		AC	49358c1537e6523cad14b2f7e515856
13412768	12				AF	7ced35e24f5c47da2b53cff7f6a27532
705095	12	65			03	9fe8eb8051b3b9e0d8ee1e39a267df4c
1331906					82	f4927f8c5a47599e0d8601961d90bf98
1041396					82	f4927f8c5a47599e0d8601961d90bf98
1255174					82	f4927f8c5a47599e0d8601961d90bf98
1335784					82	f4927f8c5a47599e0d8601961d90bf98
1097023					82	f4927f8c5a47599e0d8601961d90bf98
1339498					82	f4927f8c5a47599e0d8601961d90bf98
16738024	A6	D2			AF	1ae396879f669448e4c2c0070aabd9a7
10281691					FB	63fd8c04da8805f1730beabd70a3d0a4
210728	12				02	752294ed5cf0be7d1dfbb0f81cfd9398
23023377					DA	888458c45bce6e56705d9d0cc68730e
28022276					DA	888458c45bce6e56705d9d0cc68730e
21254481					AF	e042a80797b774eb33ba8c96d562204
15743883	12	23	47		02	1b897a6576393f792347bab63624b07

Fig. 3. The Result of Query

Table 1. The List of Query Performance

Number of Years	Number of Medicine	Executing Times(Min)
1	1	3
1	5	7
1	10	10
3	1	7
3	5	15
3	10	20
6	1	14
6	5	21
6	10	30

analyzing, we assumed that it is because we stored all data on one MySQL server, which caused a certain degree of burden to the server (average query time of a table is about 20 – 25 seconds). In conclusion, we think that if we store each year's data on different server may improve the performance.

3.2 Extended Applications

One case study was used to test our system. We created a subset results of data by a query based on the populations of patients who have been diagnosed with mouth and

neck cancer, such as oropharynx cancer, nasopharynx cancer or tongue cancer (ICD9 Code 140 – 149), and uses morphine sustainably up to three months or more. The query results of this case contained 47 patients with 351 records. Then, we processed the data into 7 variables as gender, age, duration of morphine used, dosage, whether having combine drugs, dosage of combination and change of morphine dosage. We used WEKA 3.6.0 as our analyzing tool to obtain the relationship between variables, and the result of the Bayesian Network Diagram is shown as follows, its sensitivity is 72.9%.

In Fig. 4, the Bayesian Network Diagram shows that "Change" have some connections with the other variables, for example the variables "Old-class" and "Mon". Fig. 5 shows the correlation between these three variables, such as patients who are 41 – 50 years old have the most obvious effect after using morphine for 10 – 12 months, in which 36.1% of the population will decrease morphine dosage, followed by 4 – 6 months of morphine use only 25% with decreasing dosage; For patients between 51 and 60 and over 60 years old, using morphine with 1 – 3 months is the best effectiveness, 65.6% and 53.6% will reduce the dosage, there are 32.1% of the population who is over 60 years old will reduce morphine dosage after taking 4 – 5 months. However, we still need more clinical cases to support our data.

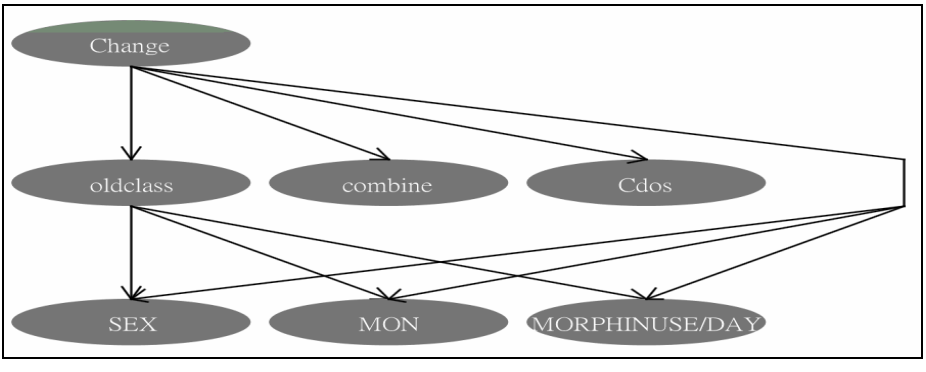


Fig. 4. Bayesian Network Diagram

Probability Distribution Table For MON

Change	oldclass	<1m	1-3m	4-6m	7-9m	10-12m	>12m
decrease	41-50	0.083	0.083	0.25	0.139	0.361	0.083
decrease	51-60	0.094	0.656	0.094	0.094	0.031	0.031
decrease	>60	0.036	0.536	0.321	0.036	0.036	0.036
decrease	<40	0.167	0.167	0.167	0.167	0.167	0.167
changeless	41-50	0.307	0.293	0.25	0.107	0.036	0.007
changeless	51-60	0.394	0.35	0.239	0.006	0.006	0.006
changeless	>60	0.462	0.388	0.112	0.012	0.012	0.012
changeless	<40	0.5	0.278	0.056	0.056	0.056	0.056
increase	41-50	0.207	0.622	0.134	0.012	0.012	0.012
increase	51-60	0.08	0.33	0.295	0.17	0.116	0.009
increase	>60	0.135	0.481	0.327	0.019	0.019	0.019
increase	<40	0.125	0.375	0.125	0.125	0.125	0.125

Fig. 5. The Relationship Between three variables

References

1. Lu, Z.-w., Lin, C.-J., Li, J.-H., Sun, W.-Z.: Taiwan morphine consumption of 16-year trend - with Asia and the world comparison, pain. Medical journal 14(1), 1–6 (2004)
2. Wan-Ping, Y.: A gold drugs or medication? Opiates is a paradox of the role, Chen-Kong university health care group
3. Taiwan Cooperative Oncology Organization, Cancer pain management guidelines. National Institutes of Health Cancer Institute (2007)
4. Cleeland, C.S., Gonin, R., Hatfield, A.K., et al.: Pain and its treatment in outpatients with metastatic cancer. N. Engl. J. Med. 330, 592–596 (1994)
5. Portenoy, R.K., Lesage, P.: Management of cancer pain. Lancet 353, 1695–1700 (1999)
6. Portenoy, R.K., Payne, D., Jacobsen, P.: Breakthrough pain: characteristics and impact in patients with cancer pain. Pain 81, 129–134 (1999)
7. Curran, D., Fossa, S., Aaronson, N., et al.: Baseline quality of life of patients with advanced prostate cancer. Eur. J. Cancer 33, 1809–1814 (1997)
8. Peng, W.L., Wu, G.J., Sun, W.Z., Chen, J.C., Huang, A.T.: Multidisciplinary management of cancer pain: a longitudinal retrospective study on a cohort of end-stage cancer patients. J. Pain Symptom Manage 32, 444–452 (2006)
9. World Health Organization, Cancer pain relief, 2nd edn. WHO, Geneva (1996)
10. de Leon-Casasola, O.A.: Current developments in opioid therapy for management of cancer pain. Clin J. Pain 24, S3–S7 (2008)
11. Benyamin, R., Trescot, A.M., Datta, S., et al.: Opioid complications and side effects. Pain Physician 11, S105–S120 (2008)
12. William, L., MacLeod, R.: Management of breakthrough pain in patients with cancer. Drugs 68, 913–924 (2008)

DICOM-Based Multi-Center Electronic Medical Records Management System to Support Clinical Diagnosis

Jui-Hung Kao[1], Chien-Yeh Hsu[2,3], and Yu-Ping Sung[2]

[1] Information System Office of National Taiwan University Hospital
[2] Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan
[3] Taipei Medical University
President, Taiwan Association for Medical Informatics
250 Wu-Hsing Street, Taipei, 110, Taiwan
cyhsu@tmu.edu.tw

Abstract. In the past, many of the medical information produced by hospitals (such as medical records or imagery) have had trouble being transferred from place to place; the main reason for this being the fference in the many formats that are used by each and every brand. With the development of the ages, the need for exchange of medical information has increased. Thus, imagery medical information uses the DICOM (Digital Image and Communication in Medicine) standard. The goal of this study is to establish a reformatting interface using the DICOM file management system to transfer paper-based records into DICOM format. Once in DICOM format, these records can then be sent to the DICOM server so that they may be accessed via the DICOM Viewer cross hospitals. This study also aims to transform text information into DICOM format so that files may be exchanged with other hospitals through DICOM's secure system so that medical records can be incorporated into hospitals' image systems to form a congregation of hospital resources.

Keywords: DICOM, Electronic Medical Records Management System, Multi-Center.

1 Introduction

For medical institutions, the use of information technology in management has become key to the modernization of hospitals in order to meet the demands that future developments might bring [2]. Environmental changes faced by hospital management arose from the establishment of various large privately owned hospitals. It would seem that the management of hospitals is, as of now, favoring expansion and dispersing of staff and resource. With each branch hospital holding its own stash of medical case files, the exchange and organization of files on a cross-branch basis has become difficult to manage.

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 130–140, 2009.

This research aims to develop a system to organize paper-based case files in the PACS stead in order to improve the quality of service. This system aims not to override the original system, but to assist in the management and of the former and any new material which may arise.

This study will focus on a setup that can digitalize files into DICOM, one that can scan files to fit DICOM's standards and be sent via the 'web not just in connection to DICOM, but also to other places to form a setup that manages the digitalization of medical files and other digital communication. Within the setup there is a medical information processor, a scanner, a DICOM code-switcher, a video transmitter, a digital video encoder and compressor and a document organizer. This entire setup will provide medical personnel with the feedback and training they may need to enhance the effects of medical information and improve the quality of medical care in general.

2 Literary Review

As medical procedures become more and more complicated with the NHS holding tighter reins on the number of beds available and overall efficiency, the present flaws in the file management system becomes obvious. First of all, preserving traditional files need sufficient labor and space; secondly, as there is often only one copy of any file, misplacement would need help from additional persons which would prevent the doctor from making a timely diagnoses and lower the overall quality of care [3]. However, a DICOM digital file would resolve both f the above problems. For example, it would make multiple simultaneous perusals of the same file possible while saving the hospital a lot of space [4]. Moreover, the imaging choices available present doctors with the ability for quicker and more accurate diagnoses.

With the popularization of the use of digital information in the field of medicine, researchers have attempted to set up systems for clinical use, including those purely for the organization of text, images and video. However, clinical use would be much elevated if data collection compatibility ceased to be a problem and data search was made possible. The two main points in this study are: (1) to transfer non-traditional medical image systems using the （Digital Imaging and Communication in Medicine, DICOM） [7][8] interface and (2) to insert search functions into various databases and explore the possibilities therein.

Since 1982, the America College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) have formed a committee to explore the connections between digital imaging equipment and computers. In 1992, the ACR/NEMA 3.0, also known as DICOM 3.0, was named. It includes the hardware needed for transmission of regular images and a data dictionary as well as commands for transmission plus a message principle, which weaves a stream of commands into messages to make convenient the communication between facilities [6]. Thus, meeting the standard of DICOM 3.0 not only allows all digital imagery to be linked via the 'web, but also for doctors to search for patient images and records to put hospital resources to better use. Moreover, DICOM 3.0 is compatible with the existing standards so that former investments need not be put to waste [5].

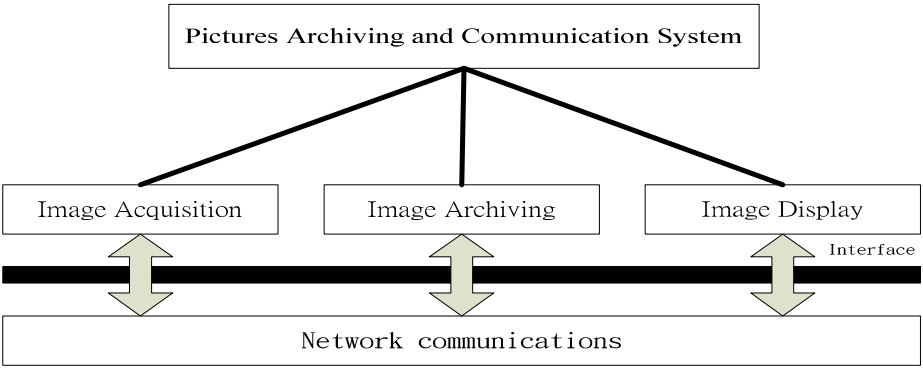


Fig. 1. Structure of medical imaging systems
(source: *Computer and Communication Magazine* Issue 53, page 41)

There are three main sub-systems to medical imaging systems: Image Acquisition, Image Archiving and Image Display [1]. These three sub-systems, through a standard interface, is connected via a computer network. Figure 1 shows that the technicalities involved in an image archive includes (1) image acquisition, (2) image processing and display, (3) image archiving and management, (4) a communication network and (5) an equipment interface and standard.

3 System Structure

Here is an example of some problems that may arise from paper-based files. Hospital A, due to different needs from different departments, set up multiple filing systems from different companies leading to a lack of consistency. When doctors wish to see cross-departmental patients, the multiple systems become a hindrance. Additional personnel are needed for system management, which also adds to financial burden. However, after installing PACS with DICOM technology, the paper-based records could be digitalized by simply running through the scanner and transferred into DICOM format through Non-DICOM Gateway digitalization. After all relevant records are in PACS, doctors may access the information safely, conveniently and consistently (as shown in Figure 2).

Using a computer workstation and a scanner connected with a USB port, test the system to see that it is working normally after installation. Afterwards, install the software needed to transfer files into DICOM format then test again and complete the setups required before testing the connection between the HIS system and the DICOM server. Open the software for DICOM format transfer, access patient information through HIS system, scan the paper records to reformat, then send all digitalized DICOM records to the DICOM server to complete for future doctors to not have to go through human processing to have access to records needed (shown in Figures 3).

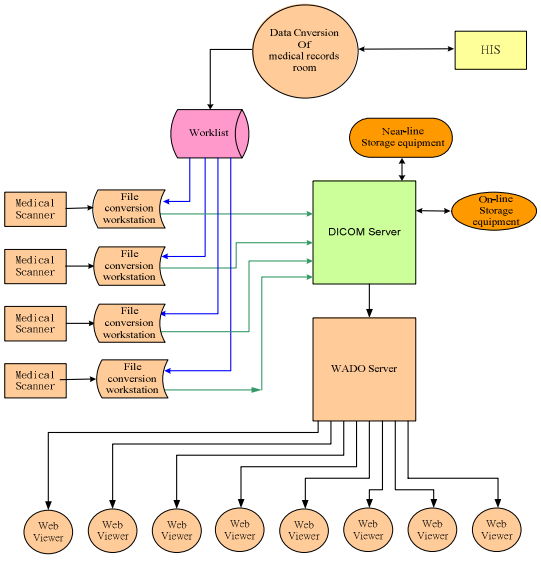


Fig. 2. Diagram of DICOM electrical records management system

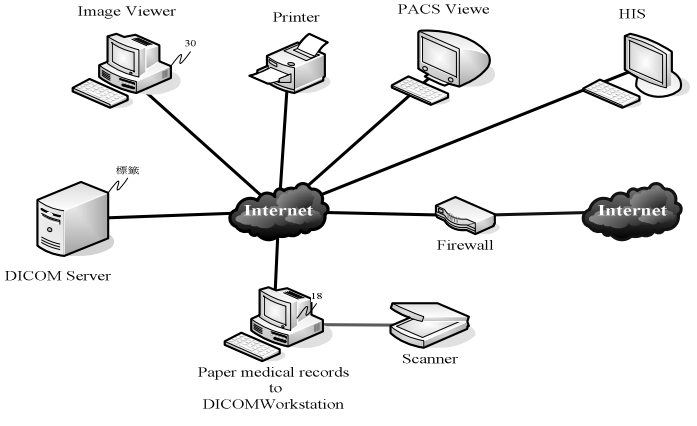


Fig. 3. Structure of DICOM electrical medical records management systems

This study provides a computer plus scanner device to turn paper files to images before going through 11, 12, 12, 14, 16, 17 to reach 20.

A filing system digitalization to DICOM device can scan paper materials into image files, then turn them into DICOM standard format to be saved onto a PACS, including:

A filing system digitalization to DICOM device can scan paper materials into image files, then turn them into DICOM standard format to be saved onto a PACS, including:

1. A medical information processor that connects with the workstation to receive commands such as file order;
2. A numbering processor to number text structurally based on information sent from the information processor;
3. A DICOM formatting device to transform text information into DICOM values after the numbering processor is connected to allow for image management;
4. A image transmitter connected to the scanner to transmit the images;
5. An image encoder and compressor that connects input from DICOM to the format values and shows a column of values to represent the information and images sent and after checking the VR and kept strings, the machine will do a thorough check of patient information before digitalizing to fit DICOM requirements. After the input of patient information, the scanner together with the encoder digitalizes the file to DICOM format before determining whether compression is needed. A final check is made to ensure consistency before editing to the patient file may be made, as shown in Figure 4;

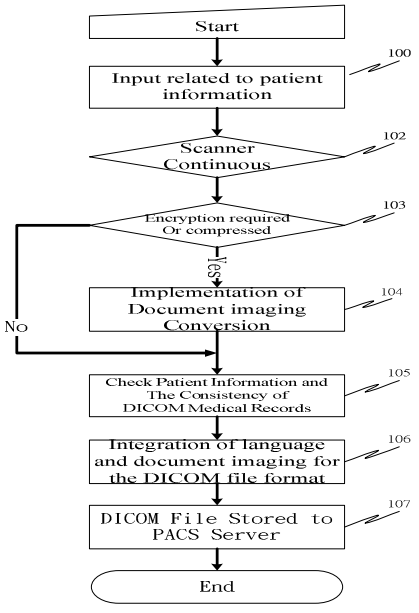


Fig. 4. Flow chart of paper-based and DICOM medical records

6. A document maker combines text and images into DICOM format and can transfer required files through broadband to a delegated DICOM server. This cycle is shown in Figure 5.

3.1 System Test Environment

As shown in Figure 6, PC-01, PC-02 and PC-03 transmitted 10, 20 and 30 DICOM medical records hrough 1000Mbps cable network. Speed of transmission was calculated through the SCU and SCP records and file size.

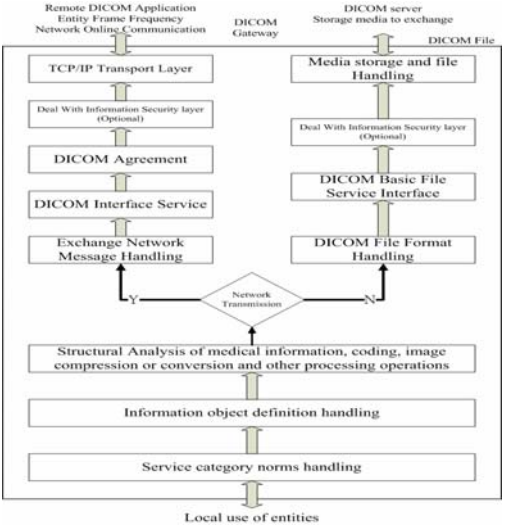


Fig. 5. Flow chart of transmission through the DICOM network

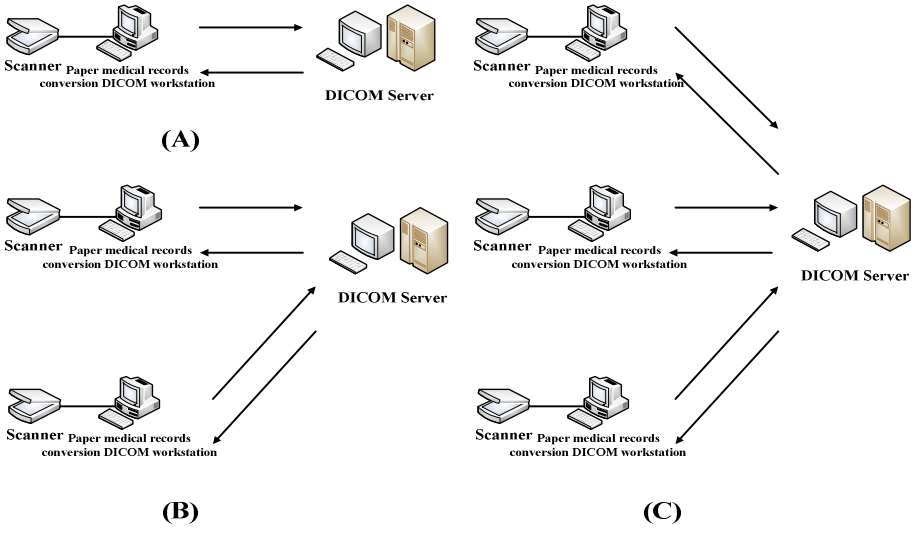


Fig. 6. Diagram of computer SCU and server SCP testing

3.2 DICOM Network Transmission

1. To Save

The DICOM network consists of two parts—the SCU (Service Class User) that makes the request, and the SCP (Service Class Provider) that receives the file. This study is designed according to DICOM standards. In principle, PACS systems are equiped for network transmissions. In Figure 7, the SCU and SCP link saving and exchange of

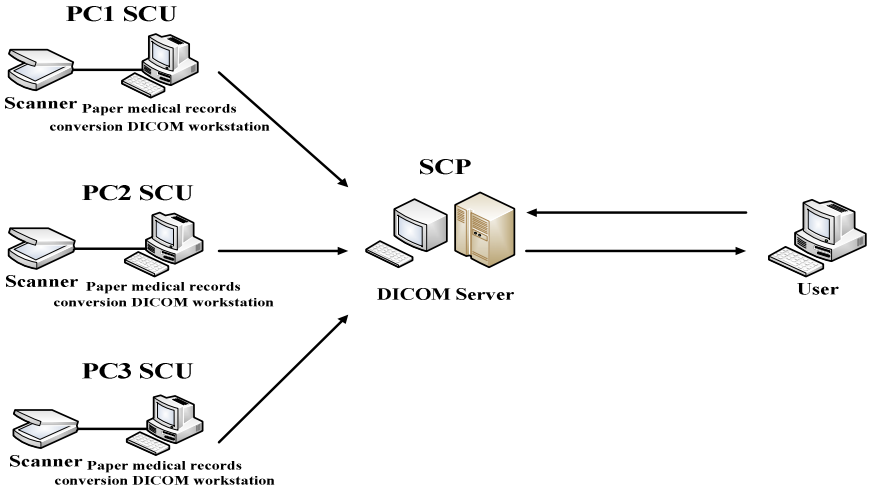


Fig. 7. Diagram of computer SCU and server SCP transmission

information of the sub-systems. As shown below, each computer is matched through SCU and SCP to transfer data to a medical digital image transmission server.

2. Query / Retrieve

The user uses workstations in the clinic or the nurses station and searches for information through the sub-systems to display the data on the workstation for clinical reference, as shown in Figure 8.

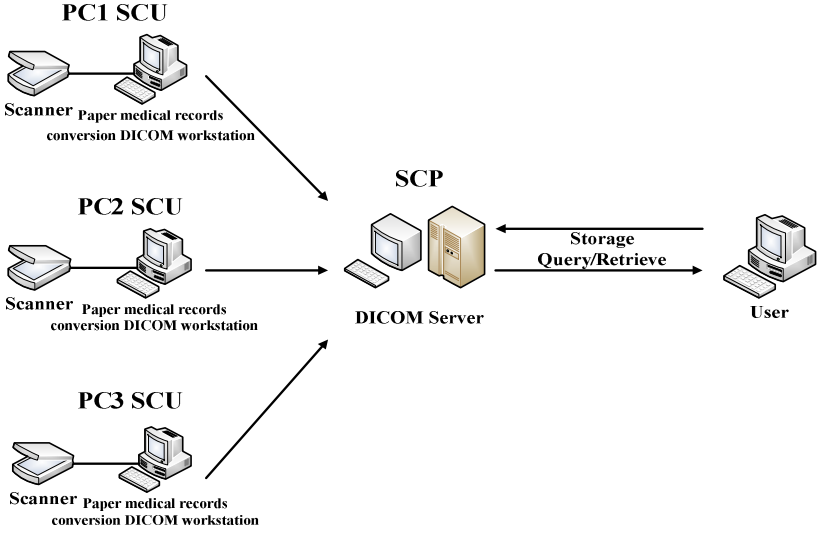


Fig. 8. Diagram of user using query / retrieve on server

4 System Test Project

DICOM Transmission Service Module Test.

4.1 Query / Retrieve Testing

As shown below, TPC1, TPC2 and TPC3 engaged in server query / retrieve through a 1000 Mbps network. Five queries were made for patient data, study data and series data, and five retrieves were made for study retrieve and series retrieve to understand system efficiency in a network environment. The structure of the server query / retrieve made is shown is Figure 9.

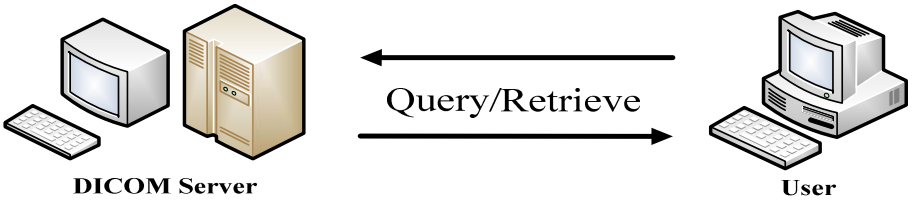


Fig. 9. Diagram of user making server query / retrieve at workstation

4.2 System Testing Results-Results of DICOM Transmission Model Testing

The transmission program splits data into SCU and the listening and receiving SCP. Before a transfer, SCU will verify the correctness of the data to be transferred before messaging with SCP to send data through the network, as shown in chart 1 and figure 10.

Chart 1. Results of transmission efficiency testing between computer SCU and server SCP

A

Exp	Image Type	Size (MB)	Time (sec)	N
1PC-User	10 Page (10MB)	10.1MB	35.3	5
	20 Page (22MB)	22.3MB	53.4	5
	30 Page (31MB)	31.2MB	194.1	5

B

Exp	Image Type	Size (MB)	Time (sec)	N
1PC-User	10 Page (10MB)	10.1MB	36.1	5
	20 Page (22MB)	22.3MB	57.7	5
	30 Page (31MB)	31.2MB	201.3	5
2PC-User	10 Page (10MB)	10.1MB	36.4	5
	20 Page (22MB)	22.3MB	58.3	5
	30 Page (31MB)	31.2MB	201.4	5

Chart 1. (*continued*)

C

Exp	Image Type	Size (MB)	Time (sec)	n
1PC-User	10 Page (10MB)	10.1MB	36.5	5
	20 Page (22MB)	22.3MB	58.7	5
	30 Page (31MB)	31.2MB	201.6	5
2PC-User	10 Page (10MB)	10.1MB	36.2	5
	20 Page (22MB)	22.3MB	57.9	5
	30 Page (31MB)	31.2MB	202.8	5
3PC-User	10 Page (10MB)	10.1MB	36.4	5
	20 Page (22MB)	22.3MB	58.5	5
	30 Page (31MB)	31.2MB	202.1	5

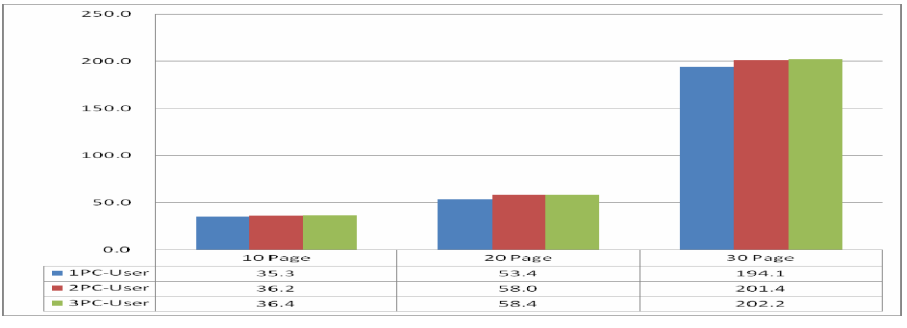


Fig. 10. Transmission time difference between multiple SCUs and SCP

He user made query / retrieve to the digitalization server to search for and retrieve the data needed. After retrieving the image from the archive, it is displayed through a sub-system at the workstation. as shown in chart 2 and figure 11.

Chart 2. Test results of query / retrieve network testing

A

Exp	Image Type	Size (MB)	Time (sec)	N
1PC-User	10 Page (10MB)	10.1MB	12.5	5
	20 Page (22MB)	22.3MB	16.8	5
	30 Page (31MB)	31.2MB	24.9	5

B

Exp	Image Type	Size (MB)	Time (sec)	N
1PC-User	10 Page (10MB)	10.1MB	13.5	5
	20 Page (22MB)	22.3MB	17.3	5
	30 Page (31MB)	31.2MB	25.0	5
2PC-User	10 Page (10MB)	10.1MB	13.3	5
	20 Page (22MB)	22.3MB	17.3	5
	30 Page (31MB)	31.2MB	25.3	5

Chart 2. (*continued*)

C

Exp	Image Type	Size (MB)	Time (sec)	N
1PC-User	10 Page (10MB)	10.1MB	13.6	5
	20 Page (22MB)	22.3MB	17.1	5
	30 Page (31MB)	31.2MB	25.6	5
2PC-User	10 Page (10MB)	10.1MB	13.9	5
	20 Page (22MB)	22.3MB	17.3	5
	30 Page (31MB)	31.2MB	25.7	5
3PC-User	10 Page (10MB)	10.1MB	13.7	5
	20 Page (22MB)	22.3MB	17.3	5
	30 Page (31MB)	31.2MB	26.5	5

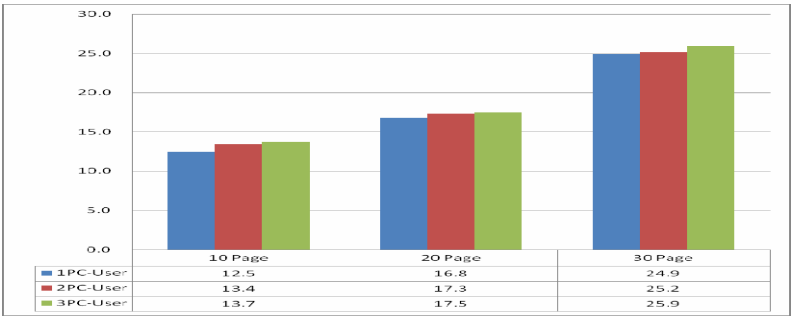


Fig. 11. Time differences between multiple query / retrieve network transmissions

5 Conclusion

Therefore, digitalizing records to fit DICOM format includes network application and file management. Its difference to other formats is that it includes all the information in the same data. DICOM files are standardized and free plus a string of images. A single DICOM document is only an image, but the image may contain many sets of images in order to save images in motion and other complex data. The images may also be compressed and used in other formats including JPEG, JPEG Lossless, JPEG 2000, LZW and Run-Length Encoding (RLE).

From the patient's point of view, a DICOM structured hospital may largely reduce waiting time. What previously may need several trips to the hospital to complete (visitation, photography then reports,) now only takes one. In a modern society that desires speed and efficiency above all else, this is a great appeal to all. For doctors, DICOM may also reduce consultation time so that doctors may have more time to engage in research. The DICOM archive also makes it easy to access data for statistic or research references.

Patient files will increase indefinitely. With speed and accuracy stressed with the retrieval of each file, a perfected computer system is needed for quality control. Thus,

the system needs to be carefully monitored and disease-sorting functions should also be subject to change and updates as needed to increase the level of management. The digitalization of medical files into DICOM images and saved onto the hospital DICOM server through coordination with PACS provide all personnel inside the hospital and out to browse medical images and intensify patient care. Mass labor for passing files around could also be saved along with storage space.

References

1. Fan, B.-y.: Health Information Management. Ho-Chi Book Publishing Co., Ltd (2006)
2. Digital Imaging and Communication Standard, ACR/NEMA Standard Publication, No.300, Virginia (1989)
3. Lead tools medical image (June 2002), http://www.leadtools.com/home2/VertMkts/ltme.htm
4. Wu, J.-S., Li, Y.-C., et al.: Medical Information Management. Wey Far Books Co., Ltd (2001)
5. Wang, C.-H.: A DICOM Medical Image Query and Access System with Browser-interfaced Functionalities, Master of Science in Electrical Engineering National Cheng Kung University (2000)
6. Chen, C.-N., Huang, S.-K., Hung, S.-W., et al.: The Experience of Accessing Filmless in a Medical Center. Chinese Journal of Radiology 29, 253–262 (2004)
7. Liang, J.-S.: Cluster-based Picture Archiving and Communication System, Graduate Institute of Medical Informatics Master Thesis, Taipei Medical University (2002)
8. Huang, S.-K.: The Installation of Performance Indicator for the PACS-The Taichung Veterans General Hospital Experience, Graduate Institute of Information Management, National Chung Cheng University (2003)

Knowledge Management on the Novel LAGE-Like GlcNAc-Transferase Protein Family

Kuo-yuan Hwa[1,2,3], Wan-Man Lin[2], Chueh-Pai Lee[2], and Mei-Yu Chen[4]

[1] Center for Biomedical Industries
[2] Departement of Molecular Science & Engineering, National Taipei University of Technology, 1, Sec 3, Chung-Hsiao E. Rd, Taipei, 116, Taiwan
[3] Institute of Medical Technology, Taipei Medical University, 250 Wu-Shin St., Taipei, 110, Taiwan
[4]Institute of Biochemistry and Molecular Biology, National Yang-Ming University School of Life Sciences, Taipei, Taiwan, ROC
kyhwa@ntut.edu.tw

Abstract. To study a novel protein family requires a biologist to understand many different types of information. Moreover, the speed of discovery in biology has been expanding exponentially. And, the public knowledge management platforms are not customerized to a research project. It is often cumbersome. To overcome these obstacles, we have set up a knowledge management platform by integrating several databases. Particularly, we have constructed a knowledge management platform on the novel LARGE-liked glycosyltransferases family. The predicted protein structure of a LARGE protein family member contains an *N*-terminal cytoplasmic domain, a transmembrane region, a coiled-coil motif and a putative catalytic domain with the conserved DXD motif and a conserved protein structural domain, as previously described. Firstly, we have developed workflow to identify the members of the LARGE-liked protein family. We then have integrated DNA sequence, protein sequence, protein domain, protein family based on the CAZY databank, human and animal disease information and references into one knowledge management platform for studying diseases.

Keywords: Protein family, Knowledge Management, LARGE-liked GlcNAcT-transferase.

1 Introduction

LARGE, was first identified within a genomic region frequently deleted in human meningioma, a type of brain tumor [1]. Moreover, mutations in the *LARGE* gene are also found in spontaneous myodystrophy *myd* mouse [2-4]. Moreover, in a patient with muscular dystrophy, LARGE genes are mutated [2, 4]. The predicted protein structure of LARGE contains an *N*-terminal cytoplasmic domain, a transmembrane region, a coiled-coil motif and two putative catalytic domains with the conserved DXD motif typical of many glycosyltransferases and a structural conserved domain as

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 141–145, 2009.

glycosyltransferase [1, 5-8]. And, the conserved features of LARGE have also been found in our previous identified novel genes [7,8]. The proximal catalytic domain is most homologous to a bacterial family GT8 glycosyltransferase of the CAZY database [9]. Members of GT8 are involved in lipooligosaccharide synthesis [10]. The distal domain have high similarity with the human UDP-GlcNAc:Gal-β1,3- N-acetytglucosaminyltransferase (iGnT) [11]. Hence, LARGE protein family is classified as members of a family of GT49 glycosyltransferase family of the CAZY database [9]. Although the glycosyltransferase activity of LARGE has not been directly demonstrated yet, evidence suggests that LARGE is required for the generation of functional, properly glycosylated forms of α-DG that can bind to laminin [12-14]. Recently a highly homologous gene *LARGE2* has been identified and cloned [15]. LARGE2 binds to α-DG and supports its maturation by glycosylation even more effectively than LARGE.

2 Classification of Eukaryote GlcNAc-Transferases

GT family is an important protein family, consists of enzymes with various biological functions. The best known examples of GT are the ones involved in cancer biology. For example, *N*-acetylglucosaminyltransferase V (GnT-V) catalyzes the addition of beta1,6-GlcNAc branching of *N*-glycans. Overexpression of GnT-V can facilitate cancer metastasis [16-18]. Another well known example is the *N*-acetylglucosaminyl-transferase III (GnT-III). GnT-III catalyzes the formation of a bisecting GlcNAc structure in N-glycans, resulting in the suppression of metastasis [19-21]. Because the sugar chains of glycoproteins are essential for the maintenance of the ordered social behavior of differentiated cells in multicellular organisms, alterations to the sugar chains are the molecular basis of abnormal social behaviors in tumor cells.

3 Construction of a Knowledge Management Platform on the LAGE-Liked GlcNAc Transferase Family

Our purposes for constructing a knowledge management platform on the LAGE-liked GlcNAc-transferases are: to collect data from public databases, on the newly discovered GnT, to assist in experiment design for functional studies, and to discover potential applications of the LARGE-liked GnT in biomedical industries. The composite knowledge manage platform includes information from the CAZY database for GlcNAc-transferase classification, the GeneBank for gene and protein sequences, the Unigene electronic northern database for studying the gene expression in human tissues, the Pubmed for literature and the OMIN database for the functional studies about the human disease, as shown in Fig 1. The source of information are as following:

CAZy (Carbohydrate- Active enZymes) database ([http://afmb.cnrs-mrs.fr/ CAZY/]),
EntrezGene([http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene]),
GenBank([http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide]),
Dictybase([http://dictybase.org/]),
Swiss-Prot([http://www.uniprot.org/]),

InterPro([http://www.ebi.ac.uk/interpro/]),
MGI([http://www.informatics.jax.org/]),
Ensembl([http://www.ensembl.org/index.html]),
HGMD([http://www.hgmd.cf.ac.uk/ac/index.php]),
UniGene([http://www.ncbi.nlm.nih.gov/unigene]),
Gene_Wiki([http://en.wikipedia.org/wiki/Gene_Wiki]),
TGDB([http://www.tumor-gene.org/TGDB/tgdb.html]),
HUGE([http://zearth.kazusa.or.jp/huge/]),
RGD([http://rgd.mcw.edu/]),
OMIM([http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim]),
CGAP([http://cgap.nci.nih.gov/]),
PubMed([http://www.ncbi.nlm.nih.gov/PubMed/]),
GO([http://www.geneontology.org/]).

The accessis available on our web-server (http://www.biomed.mse.edu.tw), upon request.

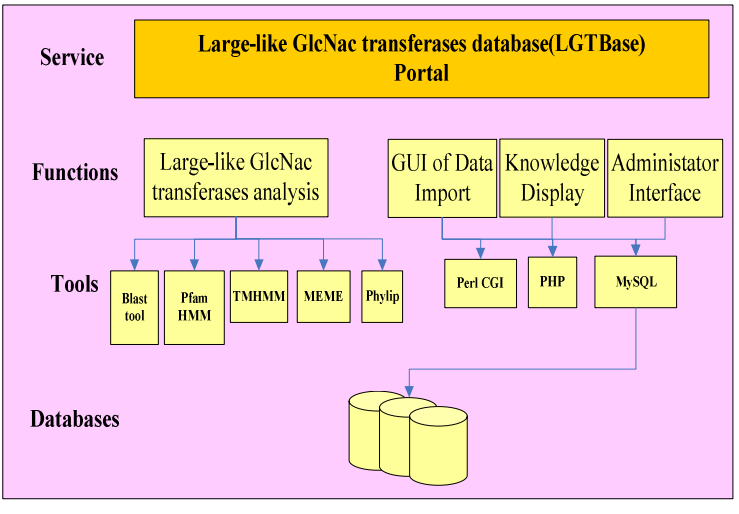


Fig. 1. Database selected for construction of the knowledge management platform

4 Utilization of the Knowledge Management Platform on the Cancer Research

By collecting the information from different databases, we are able to design our research project in related to functional studies and medical application of the novel LARGE-liked proteins identified previously [7,8] The sequences of all the LARGE-liked genes are collected. Through similarity search of sequence, five paralogs of the LARGE-liked gene are identified. The LARGE2 genes are expressed highly in human brain, prostate, kidney, and lung. LARGE1 is involved in congenital muscular dystrophy type 1D. The information can assist us to design new experiments for

collecting more new information on the funcaitonal and application of the LARGE-liked protein family. In the future, we can integrate our internal experimental data with the public data on the platform. Furthermore, the platform also can be expanded to include other protein family.

Acknowledgements. The work is supported by the grant to K.Y. Hwa from the National Science Council, Taiwan (97-2320-B-027-001-).

References

1. Peyrard, M., Seroussi, E., Sandberg-Nordqvist, A.C., Xie, Y.G., Han, F.Y., Fransson, I., Collins, J., Dunham, I., Kost-Alimova, M., Imreh, S., Dumanski, J.P.: The human LARGE gene from 22q12.3-q13.1 is a new, distinct member of the glycosyltransferase gene family. Proc. Natl. Acad. Sci. USA 96, 598–603 (1999)
2. Grewal, P.K., Holzfeind, P.J., Bittner, R.E., Hewitt, J.E.: Mutant glycosyltransferase and altered glycosylation of alpha-dystroglycan in the myodystrophy mouse. Nat. Genet. 28, 151–154 (2001)
3. Holzfeind, P.J., Grewal, P.K., Reitsamer, H.A., Kechvar, J., Lassmann, H., Hoeger, H., Hewitt, J.E., Bittner, R.E.: Skeletal, cardiac and tongue muscle pathology, defective retinal transmission, and neuronal migration defects in the Large(myd) mouse defines a natural model for glycosylation-deficient muscle - eye - brain disorders. Hum. Mol. Genet. 11, 2673–2687 (2002)
4. Longman, C., Brockington, M., Torelli, S., Jimenez-Mallebrera, C., Kennedy, C., Khalil, N., Feng, L., Saran, R.K., Voit, T., Merlini, L., Sewry, C.A., Brown, S.C., Muntoni, F.: Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of alpha-dystroglycan. Hum. Mol. Genet. 12, 2853–2861 (2003)
5. Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., Aktories, K.: A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large clostridial cytotoxins. J. Biol. Chem. 273, 19566–19572 (1998)
6. Grewal, P.K., Hewitt, J.E.: Mutation of Large, which encodes a putative glycosyltransferase, in an animal model of muscular dystrophy. Biochim. Biophys. Acta. 1573, 216–224 (2002)
7. Pang, T.L., Wu, C.J., Chen, P.A., Weng, Y.L., Chen, M.Y.: Dictyostelium gnt15 encodes a protein with similarity to LARGE and plays an essential role in development. Biochem. Biophys. Res. Commun. 360, 83–89 (2007)
8. Hwa, K.Y., Pang, T.L., Chen, M.Y.: Classification of LARGE-liked GlcNAc-Transferases of Dictyostelium discoideum by Phylogenetic Analysis. In: Frontiers in the Convergence of Bioscience and Information Technologies, pp. 289–293 (2007)
9. Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B.: An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307–317 (2003)
10. Heinrichs, D.E., Yethon, J.A., Whitfield, C.: Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enteric. Mol. Microbiol. 30, 221–232 (1998)
11. Sasaki, K., Kurata-Miura, K., Ujita, M., Angata, K., Nakagawa, K., Sekine, S., Nishi, T., Fukuda, M.: Expression cloning of cDNA encoding a human beta-1,3-N-acetyl-glucosaminyltransferase that is essential for poly-N-acetyllactosamine synthesis. Proc. Natl. Acad. Sci. USA 94, 14294–14299 (1997)

12. Kanagawa, M., Saito, F., Kunz, S., Yoshida-Moriguchi, S., Barresi, R., Kobayashi, Y.M., Muschler, J., Dumanski, J.P., Michele, D.E., Oldstone, M.B., Campbell, K.P.: Molecular recognition by LARGE is essential for expression of functional dystroglycan. Cell 117, 953–964 (2004)
13. Barresi, R., Michele, D.E., Kanagawa, M., Harper, H.A., Dovico, S.A., Satz, J.S., Moore, S.A., Zhang, W., Schachter, H., Dumanski, J.P., Cohn, R.D., Nishino, I., Campbell, K.P.: LARGE can functionally bypass alpha-dystroglycan glycosylation defects in distinct congenital muscular dystrophies. Nat. Med. 10, 696–703 (2004)
14. Patnaik, S.K., Stanley, P.: Mouse large can modify complex N- and mucin O-glycans on alpha-dystroglycan to induce laminin binding. J. Biol. Chem. 280, 20851–20859 (2005)
15. Fujimura, K.: LARGE2 facilitates the maturation of alpha-dystroglycan more effectively than LARGE. Biochem. Biophys. Res. Commun. 329, 1162–1171 (2005)
16. Przybył, M., Pocheć, E., Link-Lenczowski, P., Lityńska, A.: Beta1-6 branching of cell surface glycoproteins contribute to uveal melanoma progression by up-regulating cell motility. Mol. Vis. 26, 625–636 (2008)
17. Abbott, K.L., Aoki, K., Lim, J.M., Porterfield, M., Johnson, R., O'Regan, R.M., Wells, L., Tiemeyer, M., Pierce, M.: Targeted glycoproteomic identification of biomarkers for human breast carcinoma. J. Proteome Res. 7, 1470–1480 (2008)
18. Kim, Y.S., Hwang, S.Y., Kang, H.Y., Sohn, H., Oh, S., Kim, J.Y., Yoo, J.S., Kim, Y.H., Kim, C.H., Jeon, J.H., Lee, J.M., Kang, H.A., Miyoshi, E., Taniguchi, N., Yoo, H.S., Ko, J.H.: Functional proteomics study reveals that N-Acetylglucosaminyltransferase V reinforces the invasive/metastatic potential of colon cancer through aberrant glycosylation on tissue inhibitor of metalloproteinase-1. Mol. Cell Proteomics 7, 1–14 (2007)
19. Abbott, K.L., Nairn, A.V., Hall, E.M., Horton, M.B., McDonald, J.F., Moremen, K.W., Dinulescu, D.M., Pierce, M.: Focused glycomic analysis of the N-linked glycan biosynthetic pathway in ovarian cancer. Proteomics 8, 3210–3220 (2008)
20. Jin, X.L., Liu, H.B., Zhang, Y., Wang, B.S., Chen, H.L.: Alteration in N - acetylglucosaminyltransferase activities and glycan structure in tissue and bile glycoproteins from extrahepatic bile duct carcinoma. Glycoconj. J. 20, 399–406 (2004)
21. Guo, J.M., Zhang, X.Y., Chen, H.L., Wang, G.M., Zhang, Y.K.: Structural alterations of sugar chains in urine fibronectin from bladder cancer patients and its enzymatic mechanism. J. Cancer Res. Clin. Oncol. 127, 512–519 (2001)

Phylogenetic Analysis of HA Proteins of Influenza Virus H1N1 Reveal New Insight on Virus Pandemic

Kuo-yuan Hwa[1,2,3]

[1] Center for Biomedical Industries,
Departement of Molecular Science & Engineering, National Taipei University of Technology,
1, Sec 3, Chung-Hsiao E. Rd, Taipei, 116, Taiwan
[2] Institute of Medical Technology, Taipei Medical University,
250 Wu-Shin St., Taipei, 110, Taiwan
[3] Institute of Biochemistry and Molecular Biology, National Yang-Ming University School of Life Sciences, Taipei, Taiwan, ROC
kyhwa@ntut.edu.tw

Abstract. The pandemic (H1N1) 2009 virus is life-threatening for some patients. Analysis of virus genome sequences has provided biomedical scientists new insight in understanding host-virus interaction. We have addressed the questions if HA protein is in related to virulence of viruses. To test the hypothesis, we have focuses our analysis on the protein of hemagglutinant (HA), a protein , from the virus isolated three regions with similar living standard, popuin HongKong (death rate: 2.57 per million), Taiwan (death rate: 0.74 per million) and Japan (death rate: 0.14 per million). Our results suggest that in the human population, the virus H1N1 have isolated several time both in Taiwan and Japan, as early as in 2002 and with a dominant occurrence in 2009. However, through phylogenetic analysis, the The recent outbreak of H1N1 has different sequence origin than the previous ones. And, between the years of 2002-2008, the dominant virus stain is H3N2. The results suggest that the general population in Taiwan might not have the immunity against the new virus.

Keywords: Death Rate, Neuramidase, Phylogentic Analysis, Protein Sequence.

1 Introduction

Influenza viruses are negative strand RNA viruses. Often the viruses circulate in humans in yearly epidemics, mostly in the winter. From time to time, antigenically novel virus strains emerge sporadically as pandemic viruses [1]. Pandemic influenza viruses have emerged three times in this century. In 1918 the “Spanish” influenza (H1N1) caused approximately 675,000 total deaths in the United States [2]. Moreover, the virus killed an estimated 40 million people worldwide [3, 4, 5]. In 1957, the “Asian” influenza (H2N2) caused more than 66,000 deaths in the United States [6]. In 1968, the “Hong Kong” influenza (H3N2) [1, 7].

D. Ślęzak et al. (Eds.): BSBT 2009, CCIS 57, pp. 146–150, 2009.

2 HA Antigenic Shift and Viral Virulence

Wild water fowl is the predominant natural reservoir of influenza viruses [8]. Reassortment occurs when genetic material from avian virus strains is transferred to human virus strains. Moreover, various reassortants have been isolated from pig. They have been proposed as an intermediary host in this process [9, 10]. In 1957 and 1968, surface and internal proteins of avian virus were transferred to human influenza virus strains. The cross-species gene transfers were responsible for the pandemic influenza outbreaks [11,12]. During reassortment, change in the hemagglutinin subtype or the hemagglutinin and the neuraminidase subtype is referred to as "antigenic shift".

Recombinants Until recently there was only limited evidence that a wholly avian influenza virus could directly infect humans, but in 1997 eighteen people were infected with avian H5N1 influenza viruses in Hong Kong, and six died of complications after infection [9,10, 13,14]. Although these viruses were very poorly transmissible or non-transmissible [9, 13, 14, 15] , their isolation from infected patients indicates that humans can be infected with wholly avian influenza virus strains.

3 Phylogenetic Analysis of HA Reveal Correlation between Chronic Infection and Low Death Rate

By using the web-based service provide by Nation Center for Bioinformatic (http://www.ncbi.nlm.nih.gov/genomes/FLU/SwineFlu.html). Firstly, to test if antigenic shift is seen with the recent emergent human virus (H1N1), we have collected all the protein sequences in the database, from three regions/countries: Hong Kong, Taiwan, and Japan. The reason that the three countries were selected for analysis was because all three regions/countries had similar high quality for medical care systems. Moreover, these

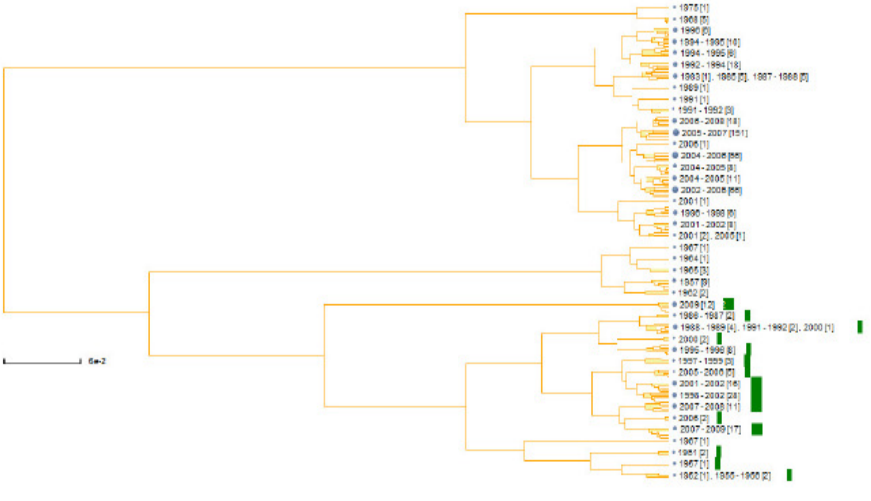


Fig. 1. Phylogenetic Analysis of HA proteins sequence from virus found in Japan. Viral sequences from H1N1 are marked in green.

three countries were nearby in Asia. Our second question was to exam the time periods for H1N1 occurrence in each region/country. Phylogenetic analyses of the HA were performed. The results from the *in silico* analysis were analyzed further with the death rate, published by the Center for Disease Control-Taiwan. The death rate for H1N1 in Japane was the lowest, with 0.14 per million people; the one for Taiwan was 0.74; and the one for Hong Kong was 2.6 per million people. The phlogenetic analysese are shown in Fig 1 (for Japan), in Fig 2 (for Taiwan) and in Fig 3 (for Hong Kong).

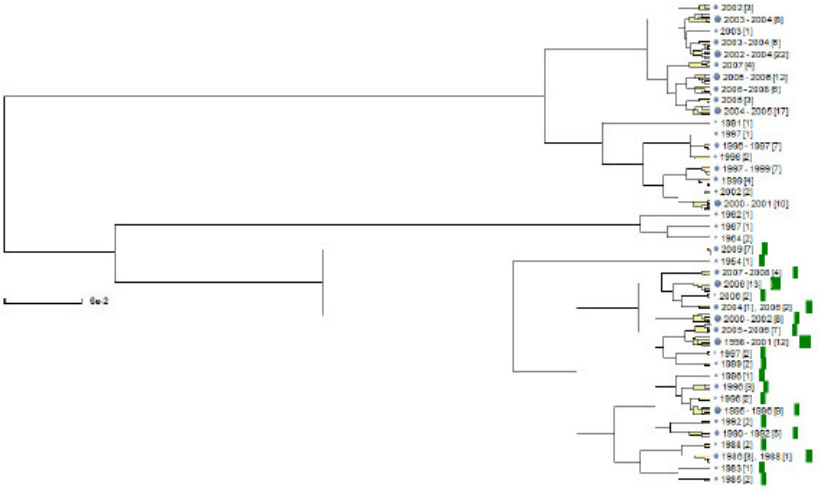


Fig. 2. Phylogenetic Analysis of HA proteins sequence from virus found in Taiwan. Viral sequences from H1N1 are marked in green.

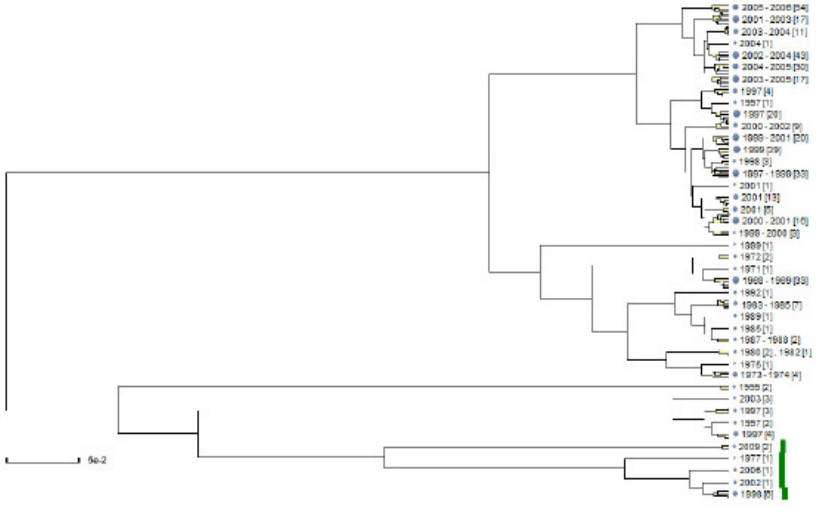


Fig. 3. Phylogenetic Analysis of HA proteins sequence from virus found in Hong Kong. Viral sequences from H1N1 are marked in green.

The new H1N1 virus, isolated in 2009, was distinct from all the previous found H1N1 viruses isolated in the three regions/countries. The results suggests that the HA sequence in 2009 influenza virus A (H1N1) was unlikely generated from mutations of previous human virus. The most likely explanation was the new HA sequences was resulted from a reassortment. And, an antigenic shift had occurred. Interestingly, when comparing with the time frames of H1N1 viruses between different regions/ countries. It is noticeable that H1N1 viruses occurred in Hong Kong only briefly. And, non-H1N1 viruses were found commonly through the years since about 1970's. However, both in Japan and in Taiwan, H1N1 viruses were commonly found through the years, as early as in about 1950's. The frequent occurrence of H1N1 in the general populations in Taiwan and Japan might give more protective immunity. Therefore, the death rates of the 2009 influenza virus A (H1N1) were compared. Indeed, the death rates were lower in Japan (0.14 per million) and in Taiwan (0.74 per million), comparing to the death rate in Hong Kong (2.6 per million). It was not surprised to found the observed correlation since HA was the molecule often used for generating protective vaccine. Similar analyses were also performed for NA. And, little correlation was found between the phylogenetic analyses and the death rate. The founding could be applicable for public control of H1N1. And, vaccination program should be helpful in Hong Kong since the phylogenetic analysis suggested few infection of similar H1N1 viruses in the region. To further confirm the finding in this study is to screen the immunological response in the general populations of the three regions/countries.

References

1. Cox, N.J., Subbarao, K.: Global epidemiology of influenza: past and present. Annu. Rev. Med. 51, 407–421 (2000)
2. United States Department of Commerce. Historical statistics of the United States: Colonial times to 1970. Government Printing Office, Washington, DC (1976)
3. Crosby, A.: America's forgotten pandemic. Cambridge University Press, Cambridge (1989)
4. Johnson, N.P., Mueller, J.: Updating the accounts: global mortality of the 1918–1920 "Spanish" influenza pandemic. Bull. Hist. Med. 76, 105–115 (2002)
5. Patterson, K.D., Pyle, G.: The geography and mortality of the 1918 influenza pandemic. Bull. Hist. Med. 65, 4–21 (1991)
6. Simonsen, L., Clarke, M.J., Schonberger, L.B., Arden, N.H., Cox, N.J., Fukuda, K.: Pandemic versus epidemic influenza mortality: A pattern of changing age distribution. J. Infect. Dis. 178, 53–60 (1998)
7. Webby, R.J., Webster, R.: Are we ready for pandemic influenza? Science 302, 1519–1522 (2003)
8. Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y.: Evolution and ecology of influenza A viruses. Microbiol. Rev. 56, 152–179 (1992)
9. Ludwig, S., Stitz, L., Planz, O., Van, H., Fitch, W., Scholtissek, C.: European swine virus as a possible source for the next influenza pandemic? Virology 212, 551–561 (1995)
10. Scholtissek, C.: Source for influenza pandemics. Eur. J. Epidemiol. 10, 455–458 (1994)
11. Kawaoka, Y., Krauss, S., Webster, R.: Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J. Virol. 63, 4603–4608 (1989)

12. Scholtissek, C., Koennecke, I., Rott, R.: Host range recombinants of fowl plague (influenza A) virus. Virology 91, 79–85 (1978)
13. Claas, E.C., Osterhaus, A.D., van Beek, R., De Jong, J.C., Rimmelzwaan, G.F., Senne, D.A., Krauss, S., Shortridge, K.F., Webster, R.G.: Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Lancet 351, 472–477 (1998)
14. Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C., Huang, J., Hemphill, M., Rowe, T., Shaw, M., Xu, X., Fukuda, K., Cox, N.: Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science 279, 393–396 (1998)
15. Katz, J., Lim, W., Bridges, C., Rowe, T., Hu-Primmer, J., Lu, X., Abernathy, R., Clarke, M., Conn, L., Kwong, H., Lee, M., Au, G., Ho, Y., Mak, K., Cox, N., Fukuda, K.: Antibody response in individuals infected with avian influenza A (H5N1) viruses and detection of anti-H5 antibody among household and social contacts. J. Infect. Dis. 180, 1763–1770 (1999)

Author Index