The Classification Performance of Binomial Logistic Regression Based on Classical and Bayesian Statistics for Screening β-Thalassemia

Patcharaporn Paokanta
Collage of Arts, Media and Technology,
Chiang Mai University,
Chiang Mai, Thailand
E-mail: patcha535@gmail.com

Napat Harnpornchai
Collage of Arts, Media and Technology,
Chiang Mai University,
Chiang Mai, Thailand
E-mail: napatresearch@gmail.com

Nopasit Chakpitak
Collage of Arts, Media and Technology,
Chiang Mai University,
Chiang Mai, Thailand
E-mail: nopasit@camt.info

Abstract—**Statistics plays an important role in many areas especially in classification tasks. Logistic Regression Model is one popular technique to solve problems, in particular, medical problems. β-Thalassemia, a common genetic disorder, lends itself to is interesting for using MLR to classify types of β-Thalassemia. There are several types of Thalassemia in the world, especially Thailand. From many methods to construct mathematical models, there are two approaches to generate these models, namely Classical and Bayesian Statistics. According to different views of both approaches, using MLR based on both approaches was selected to classify types of β-Thalassemia. The results show that classification results of all models based on Bayesian Statistics yield a greater accuracy percentage than using Classical Statistics (an accuracy percentage of this data set was 99.2126). Both approaches give different results because of the source of parameter, the transformation processes and data types are affect the classification performance based on using MLR. In the future, we will use the model most suitable for implementing Thalassemia Expert System.**

Keywords- β-Thalassemia, Classification Techniques, Binomial Logistic Regression (LR), Classical Statistics, Bayesian Statistics.

I. INTRODUCTION

In past decades, statistics has played an important role in many disciplines such as sciences, engineering, and bioinformatics [1, 2, 3]. In particular, biomedical science and biostatistics show that Classical Statistics plays a significant role in analyzing data, for example, clustering tasks, predicting tasks and correlation among others [4, 5, 6]. Along with the success of Classical Statistics in these areas, there is a mathematical modeling of data. Classical Statistics can be used as well. However, there is a longstanding and largely philosophical debate between Classical Statistics and another statistical approach, namely Bayesian Statistics. There are many differences between the two approaches. Firstly, the parameter θ is not probabilistic but fixed, or a point estimate. Secondly, an estimator of this approach can only be obtained from the function of the data. Thirdly, when n is large, classical inference based on Maximum Likelihood, is unbiased. However, in real situations, it is very difficult to prepare a data set which is large enough to use Classical Statistics, for example in medical diagnosis. This is especially so for Thalassemia, where there is research in using statistics to evaluate the types of Thalassemia. The data set for some types of Thalassemia is not large enough to use Classical Statistics. According to this constraint, Bayesian Statisticians have proposed another concept to build the parametor θ where unknown θ is considered as a random variable. Furthermore, probabilistic judgements are made directly on θ. These probabilities are called subjective as they are not necessarily based on study data [7].

In the past 20 years, constraints caused by computer technology have limited the research and use of Bayesian Statistics. Currently, however, computer technology has advanced to allow generation of the parametor θ for a particular study which has a small sample size, not large enough to use Classical Statistics. For example in the case of clinical trials, medical diagnosis and ecology there have been several applications of Bayesian Statistics e.g. breast cancer, virus genome, and leukemia [8, 9, 10, 11]. Moreover, the comparison between Classical and Bayesian Statistics were studied for example comparison of Classical and Bayesian design on clinical trials shows that Bayesian design is better than Classical design [12]. For these reasons, this paper will compare the obtained Binomial Logistic Regression model between Classical and Bayesian Statistics for screening Thalassemia. Firstly, this paper will briefly review Thalassemia in the next section.

A. Thalassemia

Thalassemia syndrome is one of the most common genetic disorders in the world. There are three main types of Thalassemia: Thalassemia minor, Thalassemia intermedia and Thalassemia major. It is estimated 1.5% of the worldwide population is diagnosed with a minor Thalassemia called β-Thalassemia. This disorder is common in areas where malaria was once prevalent, such as Africa, the Mediterranean region, the Middle East, Southeast Asia (India, Thailand and

Indonesia), and the Far East. Southeast Asia accounts for approximately 50% of worldwide carriers, while European and American countries account for 10–15%. Thalassemia has the highest prevalence in Southeast Asia, where approximately 55 million people are carriers. The gene frequency of alpha-Thalassemia reaches 30-40% in Northern Thailand and Lao PDR, while β-Thalassemia varies between 1- 9%, and HbE, which is one type of minor Thalassemia has a frequency of 50-60% at the junction between Thailand, and Lao PDR. This high magnitude in the border regions poses public health problems in Thailand. Approximately 40% of Thai people are heterozygous carriers of these genes [13].

In Thailand, Thalassemia is currently diagnosed via a two or three step process, depending on the suspected variant of disease. This method however, is slow and relies on expert knowledge and experience as well as expensive equipment. The number of genetic variations of Thalassemia in northern Thailand is large and therefore also poses particular challenges in diagnosis. Table 1. shows the number of variants.

TABLE 1. The Thalassemia types in Thailand [14].

Severeness level of Thalassemia	**Types of Thalassemia**
Thalassemia minor	1. α-Thalassemia 2 heterozygote 2. α-Thalassemia I heterozygote 3. β-Thalassemia heterozygote 4. Hb E heterozygote 5. α-Thalassemia heterozygote + β-Thalassemia heterozygote 6. α-Thalassemia heterozygote + Hb E heterozygote 7. Homozygous Hb E 8. α-Thalassemia heterozygote + Homozygous Hb E 9. Homozygous α-thalassemia 2 10. Homozygous α-thalassemia 2 + β-Thalassemia heterozygote 1 1. Homozygous α-thalassemia 2 + Hb E heterozygote 12. Homozygous α-thalassemia 2 + Homozygous Hb E 13. Hb Constant Spring (CS) heterozygote 14. α-Thalassemia 2 + Hb CS heterozygote 15. Hb CS heterozygote + Hb E heterozygote 16. Hb CS heterozygote + Homozygous Hb E 17. Hb CS heterozygote + β-Thalassemia heterozygote
Thalassemia intermedia	1. Hb H disease (All genotypes) 2. Homozygous Hb CS 3. Hb H disease + Hb E heterozygote (AE bart's disease) 4. β-Thalassemia or Hb E (Some) 5. HPFH or αβ-Thalassemia + Hb E 6. HPFH or β-Thalassemia 7. Homozygous β-Thalassemia (Some) 8. Hb H disease + Homozygous Hb E and Hb H disease + β-Thalassemia or Hb E (EF Bart's disease)
Thalassemia major	1. Hb Bart's hydrops fetalis 2. Homozygous β-Thalassemia 3. β-Thalassemia or Hb E (Some) 4. αβ Thalassemia or β-Thalassemia

This research aims to compare the multinomial logistic model between using Classical and Bayesian Statistic on the data of thalassemia particular types to help diagnosis disease.

For the propose of this study, the paper is organized as follows; the second and third section will compare the theory of Classical and Bayesian Statistics. In the fourth and fifth section; the materials and methodology such as data collection, data filtering, classification processes, data set and methodology that were used in this paper will be illustrated. The results of each process and the discussions will be demonstrated in the sixth section and finally, we will conclude the comparison of using Classical and Bayesian Statistics to classify the types of β-Thalassemia.

II. CLASSICAL STATISTICS

The foundations of Classical Statistics or conventional or frequentist were not really laid until the first half of the twentieth century. It views data as the observed realizations of stochastic systems that contain one or several random processes. However, in classical statistics, the quantities used to describe these random processes (parameters) are fixed and the constants are unknown. Moreover, in classical statistics, uncertainty is evaluated and described in terms of the

frequency of hypothetical replicates, although these inferences typically only describe knowledge from a single data set. In other words, probability in Classical Statistics is the relative frequency of a feature of observed data. Under the classical view of statistics, such statements are impossible in principle because parameters are fixed and only the data are random. In much of classical statistics, the likelihood function is used as a basis for inference. The likelihood function is the same as the sampling distribution of the observed data θ, but "read in the opposite direction": That value θ, which yields the maximum of the likelihood function for the observed data x is taken as the best estimate for θ and is called the maximum likelihood estimate (MLE) of the parameter θ. That is, much of classical inference is based on the estimation of a single point that corresponds to the maximum of a function. Note that θ can be a scalar or a vector [7].

Bayesian Statistics will be reviewed in terms of theory and discussion considering the different view between Classical and Bayesian Statistics.

III. BAYESIAN STATISTICS

Bayesian statistics is in fact very old and dominated the school of statistics for a long time if compared with Classical Statistics. Fundamentally, it is well known that this approach used conditional probability for inference embodied in Bayes rule were laid as early as 1763 by Thomas Bayes, an English mathematician. It views data as the observed realizations of stochastic systems that contain one or several random processes. Moreover, the uncertainty of Bayesian Statistics is evaluated using the posterior distribution of a parameter, which is the conditional probability distribution of all unknown quantities such as parameters, given the data, the model, and what we knew about these quantities before conducting the analysis. Forthermore, its probability is used to express one's uncertainty about the likely magnitude of a parameter; no hypothetical replication of the data set is required. Under Bayesian inference, we fundamentally distinguish observable quantities x from unobservable quantities θ. Observables x are the data, because unobservables θ can be statistical parameters, missing data, mismeasured data, or future outcomes of the modeled system (predictions); they are all treated as random variables, i.e., quantities that can only be determined probabilistically. Because parameters θ are random variables under the Bayesian paradigm, we can make probabilistic statements about them, e.g., say things like "there is a 50 percent chance of rain today".

The basis for Bayesian inference is Baye's rule, also called Baye's theorem, which is a simple result of conditional probability. The rule describes the relationship between the two conditional probabilities P(A|B) and P(B|A), where " | " is read as "given":

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \qquad (1)$$

This equation is an undisputed fact and can be proved from simple axioms of probability Bayes rule can be deduced from them, and the entire framework for Bayesian statistics, such as estimation, prediction, hypothesis testing, is based on just these three premises. In contrast, classical statistics lacks such an internally coherent body of theory. However, what used to be more controversial, and partly still is, is how Bayes used his rule. He used it to derive the probability of the parameters θ, given the data that is the posterior distribution

$$P(\theta| x) = \frac{P(x|\theta)*P(\theta)}{P(x)} \qquad (2)$$

We see that the posterior distribution is $P(\theta|x)$ proportional to the product of the likelihood function $P(x|\theta)$ and the prior distribution of the parameter $P(\theta)$. To make this product a genuine probability distribution function, with an integral equal to 1, a normalizing constant $P(x)$ is needed as a denominator; this is the probability of observing one's particular data set x. Ignoring the denominator (which is just a constant and does not involve the unknown θ) Baye's rule as applied in Bayesian statistics can be paraphrased as

Posterior distribution α Likelihood ×Prior distribution,

where α reads as "is proportional to." Thus, Bayesian inference works by using the laws of probability to combine the information about the parameter θ contained in the observed data x, as quantified in the likelihood function $P(x|\theta)$, with what is known or assumed about the parameter before the data are collected or analyzed, i.e., the prior distribution $P(\theta)$. This results in a rigorous mathematical statement about the probability of parameter θ, given the data, the posterior distribution $P(\theta|x)$. Hence, while classical statistics works by estimating a single point for a parameter (which is an unknown constant), Bayesian statistics makes inference about an entire distribution instead, because parameters are random variables described by a statistical distribution.

A prior distribution does not necessarily imply a temporal priority; instead, it simply represents a specific assumption about a model parameter. Bayes rule tells us how to combine such an assumption about a parameter with our current observations into a logical, quantitative conclusion. The latter is represented by the posterior distribution of the parameter.

Inference in Bayesian statistics is a simple probability calculation, and one of the things Bayesians are most proud of is the parsimony and internal logic of their framework for inference.

However, the requirement to determine a prior probability $P(\theta)$ for model parameters ("prior belief") has caused fierce opposition to the Bayesian paradigm because this was (and partly still is) seen to bring into science an unwanted subjective element. However, as we shall see, it is easy to exaggerate this issue, for several reasons.

1) Objective science or statistics is an illusion anyway: there are always decisions to be made, e.g., what questions to ask, what factor levels to study, whether to transform a response, and literally myriads more. Each one of these decisions may have an effect on the outcome of a study.

2) It is possible to use the Bayesian machinery for inference (Bayes rule and Markov chain Monte Carlo [MCMC] computing algorithms, see later) with so-called flat priors (also vague, diffuse, uninformative, minimally informative, or low-information priors). Such priors represent our ignorance about a parameter or our wish to let inference, i.e., the posterior distribution, be dominated by the observed data. Actually, this is exactly what we do throughout this paper.

3) The prior is seen by some statisticians as strength rather than a weakness of the Bayesian framework: it lets one formally examine the effect on one's conclusions of different assumptions about the parameters. Also, anybody using informative priors must say so and justify this choice. When the choice of priors is suspected to have an undue influence on the posterior distribution, it is good practice to conduct a sensitivity analysis to see how much one's conclusions are changed when a different set of priors is used. Nevertheless, it is fair to say that there can be challenges involving the priors.

One possible problem is that priors are not invariant to transformation of parameters. A prior that is uninformative for θ may well be informative for a one-to-one transformation $g(\theta)$ of θ, such as $\log(\theta)$ or $1/\theta$. Hence, it is possible to introduce information into an analysis without intending to do so. Especially in complex models (and these are the ones where a Bayesian treatment and the Bayesian model fitting algorithms offer the most rewards), it is quite possible that one unknowingly introduces unwanted information by the choice of ostensibly vague priors. Hence, for more complex models, a precise analysis of priors is even more useful. Still, these challenges are not seen as insurmountable by many statisticians, and Bayesian statistics has now very much entered the mainstream of statistical science [7].

From the review on the theory of Classical and Bayesian Statistics in the previous section and this section, the materials that were used for this study object will be described in the next section.

IV. Materials and Methodology

A. Materials

Thalassemia knowledge such as diagnotic processes, Thalassemia indicators were taken from experts (medical practitioner, Biochemistry) using Diagnosis Template of CommonKADS suite which is a Knowledge Engineering process. This knowledge was used to identify variables which were collected for classifying the genotypes of patients who have β-Thalassemia that are usually silent at the clinical level. In this paper, because of the prevalence of several types of Thalassemia in the northern Thailand, the data of 105 β-Thalassemi patients were collected. The data set for this experiment shows in Table 2.

The Thalassemia indicators shown in Table 2. include 12 variables which were selected by Pearson Chi-square before using Binomial Logistic Regression by Classical Statistics and Bayesian Statistics.

In the fifth section, the methodology for generating the Binomial Logistic Regression models to classify the types of β-Thalassemia between using Classical and Bayesian Statistics is presented.

TABLE 2. The Thalassemia indicators.

Variables	**Diection**
Genotype of children	Output
Types	Input
F-cell of children	Input
HbA2 of children	Input
Inclusion Body of children	Input
Genotype of father	Input
F-cell of father	Input
HbA2 of father	Input
Inclusion Body of father	Input
Genotype of mother	Input
F-cell of mother	Input
HbA2 of mother	Input
Inclusion Body of mother	Input

B. Mathodology

The methodology of this research is illustrated in Figure 1. This methodology shows that the first steps of this research is to interview experts about Thalassemia disease such as Thalassemia screening processes, types of Thalassemia using CommonKADs Templates. Moreover, the Thalassemia documents such as Thalassemia screening processes, types of Thalassemia were studied to compare the use of Classical and Bayesian Statistics to classify β-Thalassemia. In the second step, variables were defined and collected from some experts (biochemist and medical practitioner) and some documents (papers and articles from Thalassemia foundation of Thailand and Out-Patient Department records from a hospital in the northern Thailand) in the first step. These variables were elicited by Diagnosis template of CommonKads model suite. In the third step, some variables of obtained data from the previous steps were filtered using Pearson Chi-Square. Lastly, filtered variables were classified using Binomial Logistic Regression by Classical Statistics and Bayesian Statistics and finally the best model was implemented to construct an expert system which is the Knowledge Based Diagnosis Decision Support System for screening Thalassemia. This step will be developed in the future.

In the next section, the results of each step Figure 1. will be presented in term of Binomial Logistic Regression models and classification results using Classical and Bayesian Statistics.

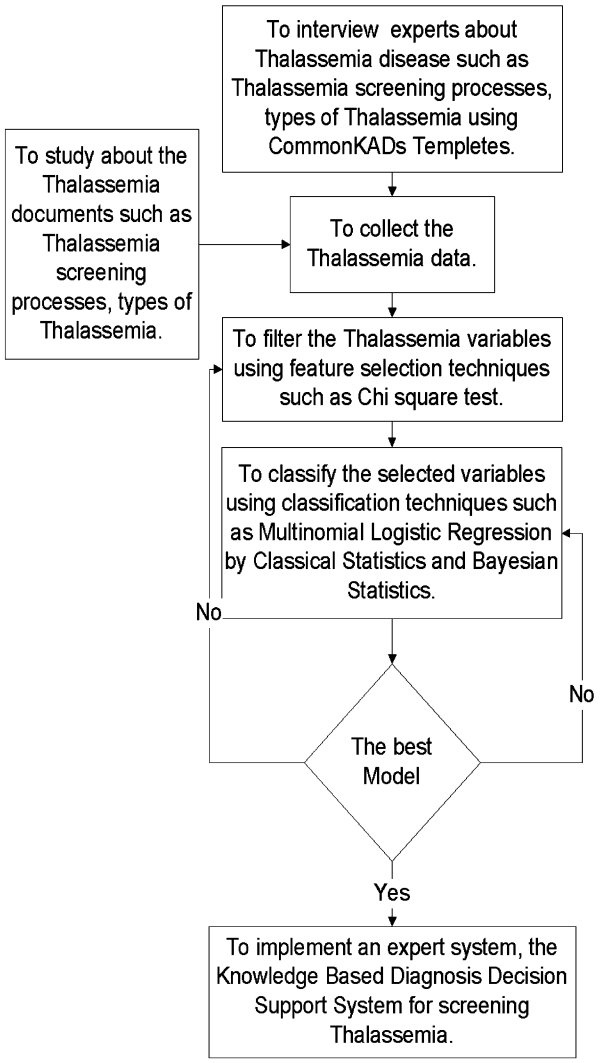

Figure 1. The methodology for classifying β − Thalassemia.

V. RESULTS

From the methodology in the previous section, the result of the first step is the type of Thalassemia that shows in Table 1. that was elicited from experts and documents by using CommonKADs. For the result of the second step, the thalassemia indicators were defined and selected as Table 2. Moreover, the result of using Pearson Chi-square in the third step is the selected variables shown in Figure 2 below.

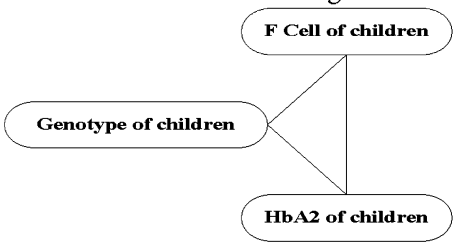

Figure 2. The Influence Network of β-Thalassemia screening indicators.

The relationship among the three variables which are Genotype, F-cell and HbA2 of children in Figure 2., two variables were inputed to classify genotype of children which can identify the type of β-Thalassemia. The fourth step classified filtered variables by using Binomial Logistic Regression Model based on Classical and Bayesian Statistics. 1

The classification performance of model obtained using Classical and Bayesian Statistics are demonstrated in the table below.

TABLE 3. The Classification results of Binomial Logistic Regression of using Classical Statistics

Model	Correctly Classified Instances (Percent)	Incorrectly Classified Instances (Percent)	Mean absolute error
(1)	114 (89.7638)	13 (10.2362)	0.1517
(2)	115 (90.5512)	12 (9.4488)	0.1123
(3)	126 (99.2126)	1 (0.7874)	0.0173
(4)	125 (98.4252)	2 (1.5748)	0.0222
(5)	124 (97.6378)	3 (2.3622)	0.0382
(6)	122 (96.0630)	5 (3.9370)	0.0394

TABLE 4. The Classification results of Binomial Logistic Regression of using Bayesian Statistics

Model	Correctly Classified Instances (Percent)	Incorrectly Classified Instances (Percent)	Mean absolute error
(7)	115 (90.5512)	12 (9.4488)	0.0945
(8)	117 (92.1260)	10 (7.8740)	0.0787
(9)	126 (99.2126)	1 (0.7874)	0.0079
(10)	126 (99.2126)	1 (0.7874)	0.0079
(11)	125 (98.4252)	2 (1.5748)	0.0157
(12)	123 (96.8504)	4 (3.1496)	0.0315

Table 3. and Table 4. show that the comparison of accuracy percentage of Binomial Logistic Regression models between using Classical and Bayesian Statistics. For type 1 of β-Thalassemia model (1) which is the obtained model of using Classical Statistics, an accuracy is 89.7638. On the other hand the obtained model of using Bayesian Statistics is model (7) which shows 90.5512 accuracy. This demonstrated that

Bayesian model yields a better result than the Classical model.

In addition, in the case of type 2 of β-Thalassemia, an accuracy percentage of model (2) and model (8) are 90.5512, 92.1260 repectively. For this type of β-Thalassemia the results show that the Bayesian model yields a better result than the Classical model.

Also, in the case of type 3 of β-Thalassemia, model (3) yields 99.2126 (the same accuracy percentage as model (9)). This implies that the classification performance of Classical and Bayesian model are not different. In case of type 4, model (4) and model (10) present that an accuracy percentages are 98.4252, 99.2126 repectively. For this case, we see that the Bayesian model gives a better classification performance than the Classical model. For the result of model (5) and model (11), their accuracy percentages are 97.6378 and 98.4252. Here, the implies that Bayesian model results are better than the Classical model. Lastly, an accuracy percentage in case of type 5 of β-Thalassemia which are model (6) and model (12), are 96.0630 and 96.8504 repectively.

From the classification results in Table 3. and Table 4., it is clear that the Bayesian model yields a greater accuracy percentage better than the Classical model in all models for classifying type of β-Thalassemia with the same Mean Absolute Error.

VI. Conclusion

The goal of this study was to compare the classification performance of constructed model using Classical and Bayesian Statistics for screening types of β-Thalassemia, which is demonstrated using Binomial Logistic Regression and represented in term of accuracy percentage. One can observe that the Bayesian model better results than the Classical model. In the previous paper [13, 15], the classification results show that the best results of using Pearson Chi-square and PCA to filter variables before using serveral algorithms for classifying types of β-Thalassemia, are K-Nearest Neighbors (KNN) and Multi-Layer perceltron (MLP) with an accuracy percentage 88.9800 and 86.6142 respectively. The transformation processes which were used on the data set improved the quality of data and lead to a better classication performance. For all results, One can say that data types of a variable play an important role for increasing classification performance.

As a result of this study and all obtained results, the obtimum models will be selected to implement the Thalassemia Expert System in the future (See Table. 4).

ACKNOWLEDGEMENT

The authors would like to thank Dr. Frank Bozel for improving the English in this paper and we appreciate his kind support.

References

[1] Luisa Turrin Fernholz, Stephan Morgenthaler and Werner Stahel. "Statistics in Genetics and in the Environmental Sciences." IEEE Engineering in Medicine and Biology, vol. 21, pp. 159, 2002.

[2] Carl Taswell, "Knowledge Engineering for PharmacoGenomic Molecular Imaging of the Brain" in *Proc.* Fifth International Conference on Semantics, Knowledge and Grid, 2009, pp.26-33.

[3] Muhammad Shoaib.B.Sehgal, Iqbal Gondal and Laurence Dooley, "Statistical Neural Networks and Support Vector Machine for the Classification of Genetic Mutations in Ovarian Cancer" in *Proc.* IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2004, pp. 140-146.

[4] Razan Paul, Abu Sayed Md and Latiful Hoque, "Clustering Medical Data to Predict the Likelihood of Diseases" in *Proc.* Fifth International Conference on Digital Information Management, 2010, pp.44-49.

[5] Yanwei Xing, Jie Wang, Zhihong Zhao and Yonghong Gao, "Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease" in *Proc.* International Conference on Convergence Information Technology, 2007, pp.868-872.

[6] Jose´ L. Sevilla, Vı´ctor Segura, Adam Podhorski, Elizabeth Guruceaga, Jose´ M. Mato, Luis A. Martı´nez-Cruz, Fernando J. Corrales and Angel Rubio, "Correlation between Gene Expression and GO Semantic Similarity" IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 2, No. 4, pp.868-872, 2005.

[7] Marc Kéry. "Introduction to the Bayesian Analysis of a Statistical Model" in *Introduction to WinBUGS for ecologists: A Bayesian approach to regression, ANOVA, mixed models and related analyses*, 1st ed., Ed. London: Elsevier, 2010, pp. 25-29.

[8] Nicola J. Cooper, Keith R. Abrams, Alex J. Sutton, David Turner and Paul C. Lambert, "A Bayesian Approach to Markov Modelling in Cost-Effectiveness Analyses: Application toTaxane Use in Advanced Breast Cancer" Journal of the Royal Statistical Society. Series A (Statistics in Society), vol. 166, No. 3, pp.389-405, 2003.

[9] X. Zhou, X. Wang and E.R. Dougherty, "Multi-class cancer classification using multinomial probit regression with Bayesian gene selection" in *Proc.* IEE System Biology, 2006, pp.70-78.

[10] I-Shou Chang, Chi-Chung Wen, Yuh-Jenn Wu, Pramod K. Gupta, Shih Sheng Jiang, Jyh-Lyh Juang and Chao A. Hsiung, "A Bayesian Approach to Virus-Gene Expression Time Course Data" in *Proc.* Emerging Information Technology Conference, 2005, pp.1-2.

[11] Ho Sun Shon, Kenu Ho Ryu and Kyung-Sook Yang, "Feature Selection of Gene Expression Data Using Regression Model" in *Proc.* 10th IEEE International Conference on Computer and Information Technology, 2010, pp.1442-1447.

[12] Roger J. Lewis and Donald A. Berry, "Group Sequential Clinical Trials: A Classical Evaluation of Bayesian Decision-Theoretic Designs" Journal of the American Statistical Association, vol. 89, No. 428, pp.1528-1534, 1994.

[13] Patcharaporn Paokanta, Napat Harnpornchai, Somdat Srichairatanakool, and Michele Ceccarelli, "The Comparison of Classification Performance of Machine Learning Techniques Using Principal Components Analysis: PCA for Screening β-Thalassemia" in *Proc.* 4th IEEE International Conference on Computer Science and Information Technology, 2011, pp. 316-319.

[14] Patcharaporn Paokanta, Michele Ceccarelli and Somdat Srichairatanakool, "The Effeciency of Data Types for Classification Performance of Machine Learning Techniques for Screening β-Thalassemia," in *Proc.* ISABEL, 2010, pp. 1-4.

[15] Sutat Fuljalen and Prawet Wasee. "Thalassemia Clinical Manifestation and Pathophysiology" http://www.mb.mahidol.ac.th/thalassemia/ modules/ news/index.php?storytopic=4&start=70, Apr. 4, 2011. [Apr. 4, 2011].