

## Decision Factors on Effective Liver Patient Data Prediction

Hoon Jin<sup>1</sup>, Seoungcheon Kim<sup>2</sup> and Jinhong Kim<sup>3,\*\*</sup>

<sup>1</sup>*Dept. of Computer Engineering, Sungkyunkwan University, South Korea*

<sup>2</sup>*Dept. of Information Network, Hansung University, South Korea*

<sup>3</sup>*Dept. of Computer Engineering, Hansung University, South Korea*

<sup>1</sup>*bioagent@skku.edu, <sup>2</sup>kimsc@hansung.ac.kr, <sup>3</sup>jinhkm@gmail.com*

### Abstract

*Various types of stress and irregular eating habits, as well as inhalation of alcohol and ongoing toxic gas, ingestion of contaminated food, excessive consumption of pickled food and drug intake, enables liver disease patients to grow up year by year. To this end, variety of data mining algorithms can help medical doctors in diagnostics of patients at the hospital. This paper treats an evaluation of the analyzed results of classification algorithms selected for better prediction based on the characteristics of data from the data set with liver disease. We investigated and analyzed the classification algorithms such as Naïve Bayes, Decision Tree, Multi-Layer Perceptron and k-NN used in a previous study, which developed our data set, and additionally Random forest, Logistic which proposed by us. Those algorithms were compared in several kinds of evaluating criteria like precision, recall, sensitivity, specificity, and so on. Through the experiments, we could know that in view of precision, Naïve Bayes is preferable than others, but in other criteria such as Recall and Sensitivity, Logistic and Random Forest took precedence over other algorithms in the performance of prediction test as considering the algorithmic characteristics to liver patient data set.*

**Keywords:** *Decision factor; Choice; Algorithms; Evaluation Criteria; Random Forest; Logistic regression; WEKA; Liver Patient Data*

### 1. Introduction

There have been many changes in our lives which like food, clothing and shelter as more and more complex and evolved the society was. As a result, number of difficulties emerged for the medical expert who specialized in treatment of disease from diagnosing and treating patients according as subspecies of the disease were come out that had not occurred frequently before as well as a variety of new diseases which were previously unseen arisen. Recently, data mining techniques in the field of artificial intelligence began to be used in clinical treatment process so that clinical data mining process had become an essential technique to be utilized in diagnosing and treating patients by physicians. However indiscriminative abuse of data mining algorithms may cause critical hazard in diagnosis. Generally the effectiveness of data mining can be varied greatly corresponding to which materials you analyze, even which method you use despite of the same data used. Therefore it is desirable for him to learn the process about how the data can be suitably treated on which evaluating criteria after understanding many useful algorithms in data mining. In this paper we demonstrated several experiments using properly proposed data mining algorithms for prediction test and compared them on some evaluating criteria to the data set containing liver

---

\*\* Corresponding author is Prof. Jinhong Kim

disease patients collected from Andhra Pradesh region of India. For the experiments, total 6 numbers of classification algorithms are used, which are comprised of our suggested 2 ones and 4 ones used in the previous study [1]. We introduce about the data and the methods in next section and explain our deep understanding about the algorithms and evaluating criteria in Section 3. In Section 4 and 5, we describe experimental plans and prediction test results on evaluation criteria. Lastly we will end up the paper as allowing you to know why this study is a valuable for effective data mining in medical diagnosis in the conclusion.

## 2. Materials & Methods

This section introduces about our data and classification algorithms used for prediction tests. As aforementioned, we focus on prediction of the clinical data, i.e., in diagnosing and treating patients with liver disease by utilizing data mining algorithms opportunely. We describe 4 algorithms that are Decision Tree, k-Nearest Neighbor, Multi-Layer Perceptron and Naïve Bayes. Also we suggest two algorithms which expected to output better results as considering the characteristics of data distribution. So we shortly brief 9 useful performance criteria.

### 2.1. Data and Descriptions

What we used Indian liver patient data set in this paper was downloaded from UCI machine learning repository [2]. It includes 414 liver patients and 165 suspicious liver people who are not decided as patients. Also it is composed of males and females, who are 439 and 140 respectively. This data was collected from Andhra Pradesh area located in north-east of India and labeled with the class values for identifying whether they are liver patients or not.

**Table I. Data Type [3]**

<b>Attribute</b>	<b>Types (Annotations)</b>
Age	<i>Numeric</i>
Gender	<i>Nominal</i> (male, female)
Total Bilirubin	<i>Continuous</i> . As the sum value of the direct bilirubin and the indirect bilirubin, it is with a diagnosis of jaundice, metabolic processes, and used at the time of observation. (>> normal: 0.3~1.9 mg/dL)
Direct Bilirubin	<i>Continuous</i> . What is shared with other molecules, conjugated bilirubin combined with glucuronic acid combined with water-soluble molecules (>> normal: 0.0~0.4 mg/dL)
Alkaline Phosphotase	<i>Continuous</i> . As hydrolase enzyme responsible for removing phosphate groups from many types of molecules, including nucleotides, proteins, and alkaloids, it is used to diagnose hepatobiliary or bone disease in the diagnosis. (>>normal: 20~140 UL/L)
SGPT (Alanine Aminotransferase)	<i>Continuous</i> . Its full name is serum glutamic-pyruvic transaminase, ALT is found in plasma and in various bodily tissues, but is most commonly associated with the liver. (>>normal: male <=45 IU/L, female <=34 IU/L)
SGOT (Aspartate Transaminase)	<i>Continuous</i> . Its full name is serum glutamic oxaloacetic transaminase, AST is found in the liver, heart, skeletal muscle, kidneys, brain, and red blood cells, and it is commonly measured clinically as a marker for liver health. (>>normal: male 8~40 IU/L,

	female 6~34 IU/L)
Total Proteins	<i>Continuous.</i> A biochemical test for measuring the total amount of protein in blood plasma or serum. Concentrations below the reference range usually reflect low albumin concentration, for instance in liver disease or acute infection. (>>normal: 6.0~8.0 g/dl)
Albumin	<i>Continuous.</i> The most common family of globular proteins is serum albumin. Albumin is the main protein of human plasma. (>>normal: 3.5~5.0 g/dL)
A/G Ratio	<i>Continuous.</i> Albumin/globulin ratio. It provides information about the amount of albumin you have compared with globulin, a comparison called the A/G ratio. It is useful when to suspects Liver damage, Spleen problems, Thymus malfunction, Kidney disease/damage, Protein digestion and absorption, Protein intake and Autoimmune conditions. (>>normal: 1.2~1.5)
Class	<i>Nominal</i> (Patient: 1, Normal: 2)

## 2.2. Classification Algorithms

**2.2.1. Decision Tree (DT):** Decision Tree is a powerful technique that is commonly used for classification and predictive approach for decision making. In DT, there are two important steps. The first one is to make the root of the tree and the second is to make each branch of the tree for detailed conditions. The solutions for our goal are represented as the leaves of the tree. Unlike the structural analysis methods of the neural network, it describes the rules with tree structure. In binary tree, each node creates two child nodes which has an answer to the yes/no question. As for a binary tree, forming two sections based on the initial separation criteria, it further forms the tree to seek criteria which can be classified the target variable into one value. When more separation is not performed and diversity has been effectively reduced, it forms tip nodes [4].

**2.2.2. k-Nearest Neighbor (k-NN):** k-NN is one of the nearest neighboring search method. The nearest neighboring search measures the degree of similarity between the case representing the current problem and all cases in the database which is the collection of past cases in order to search for similar cases of the current problem. After computing similarity, it finds instances higher than a certain threshold [5]. Therefore, it is a method which does not intentionally find determined probability arguments (parameters) for each sample but classify it into the most similar in reference set or as belonging to the class closer in the distance by displaying the coordinate value for it directly.

**2.2.3. Multi-Layer Perceptron (MLP):** Multi-Layer Perceptron has a multi-layered neural network structure that one or more intermediate layers exist between the input and the output layer. There are, as the representative neural network algorithms, MLP, RBFN, Kohonen Features maps, Hopfield Network and ART. In general more the number of hidden layers are, more the characteristics of the decision boundary of MLP increase and sophisticate. But because it is rare that there needs of more than 2 hidden layers in most of MLP experiments finding for the predictive model, normally the performance of the MLP can be changed depending on how to design the number of neurons being exist [6].

**2.2.4. Naïve Bayes (NB):** Naïve Bayes is a document classification method based on the probabilistic theory. It selects the category having the highest probability among them after calculating the probability of any document belonged to a particular category [7]. In this point, NB has the best performance when satisfy the assumption that the probability distribution following actual data has conditional independency. However, since real world' problems doesn't keep up the assumption of Naïve Bayes, we cannot expect good performance in many cases. The representational power of NB is known when the variable has a binary value to be displayed in the same form as the linear classifiers [8].

**2.2.5. Logistic Regression (Logistic):** Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables, which are usually (but not necessarily) continuous, by using probability scores as the predicted values of the dependent variable [3]. Odds are the ratio between the probability of success over the probability of failure, that is,  $\pi / (1-\pi)$ , whereas  $p$  is the probability of an example belonging to Class 0. When  $p > 0.5$ , a sample should be classified as Class 0. Otherwise it should be judge as Class 1. Since the logistic regression outputs probabilities based on the following equation:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k \quad [9]$$

, the coefficients refer to each  $\beta_i$ . Odds ratios are simply the exponential of the weights we found before. The coefficients are in fact the weights that are applied to each attribute before adding them together. However, the result is the probability that the new instance belongs to class yes ( $> 0.5$  means yes).

**2.2.6. Random Forest (RandomF):** Random Forest classifies all trees in the forest in the classification process by combination of the prediction of the tree structure each sampled according to the same distribution and the random vector values. In other words, it is said that RF do, as an ensemble method of multiple trees, better to handle categorical data after obtaining the final solution in the majority voting system for the results of each tree is judged [10].

### 2.3. Evaluation Criteria

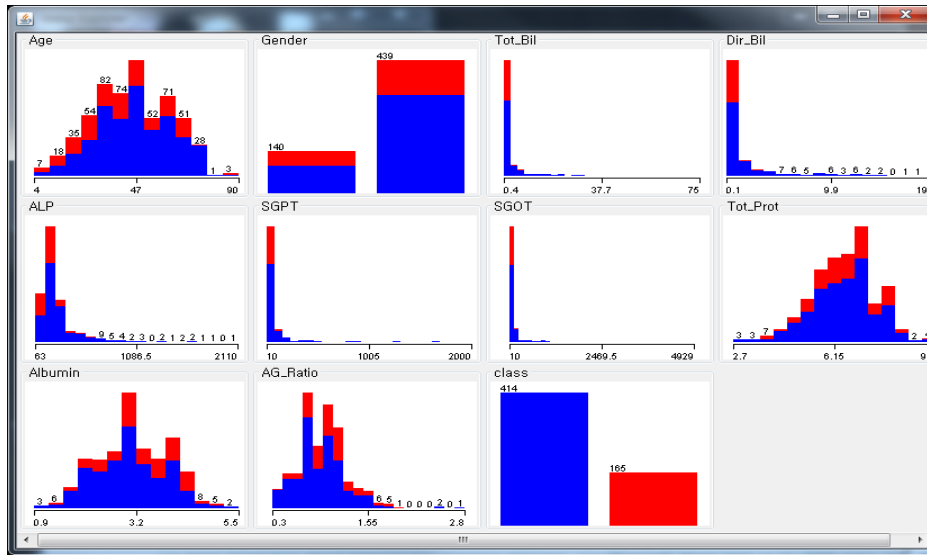
We analyze the above described six algorithms by experimental comparison. For the purpose of it we use WEKA which is most widely known as a useful tool for data mining. WEKA does not provide the environments which enables to use variety of machine learning algorithms but also supports several useful illustrated statistics [11].

- **Correctly Classified Instances (CCI):** It means how all the samples in the data set are classified correctly. Sometimes it can be interpreted as accuracy whereas Incorrect Classified Instances (ICI) may be interpreted as opposite cases to CCI.
- **Kappa statistics:** A representative statistic value measuring the degree of matching between inter-raters is Cohen's kappa. If two evaluators evaluated any item across several features as 0 or 1, Cohen's kappa value is characterized if 0-0.2, as slight, 0.2-0.4 as fair, 0.4-0.6 as moderate, 0.6-0.8 as substantial and 0.8-1.0 as almost perfect agreement [12].

- **Area under the ROC curve:** It is an enclosed area by ROC graphs which to evaluate the accuracy of detector, classifier or model including the ratio of true positive and false positive. As its shape closer to the upper left vertex of the coordinates, it has a good performance and also its size can be used to estimate the performance [13].
- **RMSE:** As the abbreviation form of the root mean square error, it denotes a frequently used measure of the difference between values predicted by a model or an estimator and the values actually observed [3]. For calculation of the RMSE, it needs of calculation of the average value of the squared errors or deviations and finally requires the root of the mean. RMSE is as the generalized formula of the standard deviation, it is used to indicate how much the difference between the estimated value and the actual value.
- **Precision/Recall:** Precision is defined as the ratio of true positive elements to all positives which are classified into the true positive and the false positive. The recall refers to the percentage of the searched relevant instances found within the retrieved data by the system.
- **Sensitivity/Specificity:** Sensitivity appears to the true positive rate which represents the ratio of the positive elements that are correctly identified. Specificity serves as the true negative rate is the proportion of negative elements that are correctly identified [1].

#### **2.4. Deep Understanding for Algorithms and Criteria**

When looking into the above 6 algorithms, we have found the followings: Generally it is known that Naïve Bayesian makes the best performance when the assumption is satisfied that the probability distribution generating the real data is conditionally independent. But we cannot be sure that the features of our data set are independent between each other because it's not fully disclosed about what are causative factors in liver disease; Decision Tree uses a greedy approach to induce trees, and in that case, it may fall to the local optima when optimizing some splitting criterion at each node, so it's difficult to reach to global optimization; Multi-Layer Perceptron is used to find the solution mainly in the problem solving process which not clearly defined in mathematics such as pattern classification, function approximation and etc. But because most of our data consist of numerical values only, relatively it can be said to be clear mathematically; k-Nearest Neighbor has an advantage of good working if input data are described structurally (means normalization) for computation between the distances. However as like Figure 1, most of features are so biased toward one side that may cause the imbalanced data problems; On the other hand, Logistic regression uses a sigmoid shape function applicable to the disease prediction. Also, it can control confounding and assess interaction very effectively in that several confounders or the confounder is a continuous variable [14]; Random Forest is expected to work appropriately in imbalanced data set because of supporting random sampling in collecting process of training samples.



**Figure 1. Data Distribution on each Feature**

Further, as we described in the above, the sensitivity measures that if there is something, something is present, and the specificity does that if there is nothing, something isn't present (see Figure 2). That is, if diagnostic criteria become to be strict, the degree of sensitivity is going down and that of specificity is going up. Therefore it is advantageous for the sensitivity to be high if needs early diagnosis though it is profitable for the specificity to be high in the area of low prevalence of disease. The accuracy means the ratio of the number of correct answers among them a classifier has found so measures the probability of not finding the incorrect answers by a classifier. Thus the accuracy describes how correct the found answers are without fault. The recall means the ratio of number of correct answers which a classifier has found among all the existing ones so measures the probability of how many correct answers a classifier has found. Resultantly, comparing in views of the value, the sensitivity equals to the recall. So the recall is important together with the sensitivity in that its values describe how many existing answers a classifier can find.

	Disease: Patient	Normal		
Positive	<p><i>True positive</i> (a)</p>	<p><i>False positive</i> (b)</p>	$Sensitivity = \frac{a}{(a + c)}$	$Precision = \frac{a}{(a + b)}$
Test: Negative	<p><i>False negative</i> (c)</p>	<p><i>True negative</i> (d)</p>		

**Figure 2. Statistical Meaning for Sensitivity, Specificity, Precision and Recall**

### 3. Experiments

#### 3.1. Previous Study Result

We referred to [1] for the comparative study. In that, after using four algorithms such as NB, DT, MLP and k-NN, they evaluated the results on the basis of 4 criteria, which are accuracy, precision, sensitivity, specificity. [1] used ranking algorithm for feature selection available in WEKA and ordered them by priority on the class. The averages of accuracy, precision, sensitivity and specificity of them are 96.552, 93.698, 0.921 and 0.986 with 12 features, respectively. But when we tried to output prediction results only with default parameters and no filters, the results were very lower than the previous study. It might be presumed that the previous study used one more feature, Globulin, and also used feature selection with ranking algorithm. Our results are showed in the Table 1.

**Table 2. A Test Result from Simulated the Previous Study**

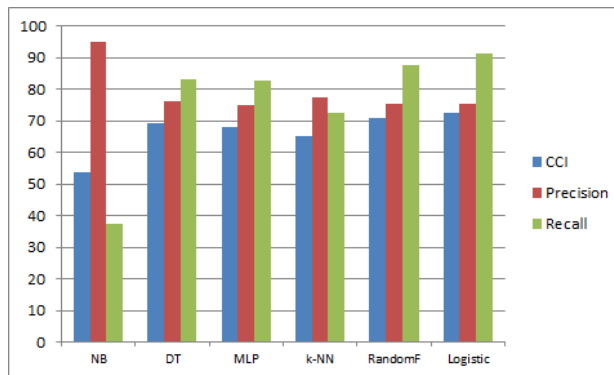
Criteria	NB	DT	MLP	k-NN
Accuracy	53.9	69.4	67.9	65.3
Precision	95.1	76.3	74.9	77.4
Sensitivity	0.374	0.831	0.829	0.727
Specificity	0.952	0.352	0.303	0.467

#### 3.2. Our Prediction Test

In this study, cross validation method was used for comparison of selected classification algorithms. It is referred also as rotation estimation, a model validation technique for evaluation of the statistical analysis generalizing to an independent data set. For the accuracy and reliability of experiments, after we repeated 10 times in each case respectively we used the average value.

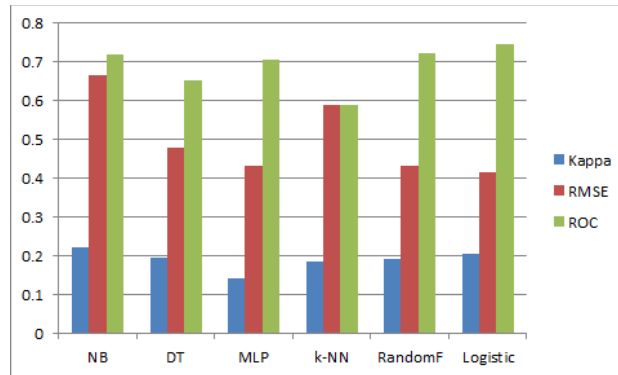
### 4. Results

Seeing Figure 3, we can see that the accuracy of Logistic shows the highest which noted as 72.7. At the recall value, that of Logistic has the highest value as 91.3, and RandomF has the second highest value in it. On the whole, the algorithms we proposed strategically make better performances at the CCI and recall relatively than the previous four algorithms which are NB, DT, MLP and k-NN.



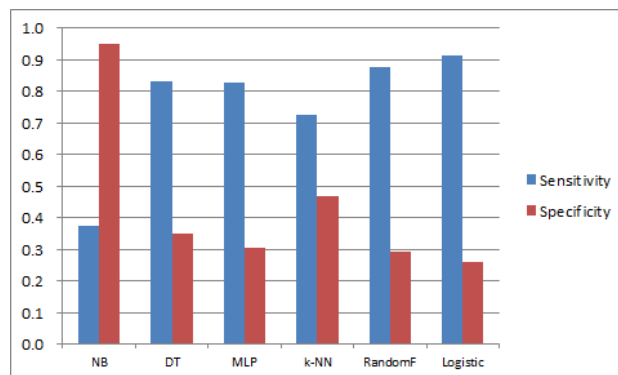
**Figure 3. Comparison of CCI, Precision and Recall**

In Figure 4, RandomF and Logistic show higher values than others at the ROC and Kappa. Also at the RMSE which describing the prediction error rate, both algorithms show to be similarly lower appearance in comparison with the previous algorithms.



**Figure 4. Comparison of Reliability and Error Rates**

Figure 5 appears to have the lowest value at the sensitivity of NB, which is 0.31. However at the specificity, NB shows the highest value and RandomF and Logistic show relatively higher than others.



**Figure 5. Comparison of Sensitivity and Specificity**

## 5. Discussion

In this study, we introduced the proposed algorithms for providing relatively higher performance in predicting patient groups by using several data mining algorithms from the data sets consisting of normal subjects and patients with liver disease compared to the algorithms used in the referred study. Naïve Bayes classifier showed higher performance than those of ours at precision as 95.1. But we could find that in the cases of CCI, namely accuracy, that representing the performance of classification beyond precision, our proposed Random Forest and Logistic showed relatively higher values than others. In the fields of science, engineering, industry, and statistics, the accuracy means the degree of closeness of measurements of a quantity to that quantity's actual value and the precision, which is sometimes called reproducibility or repeatability, is a degree to which repeated measurements under unchanged conditions show the same results [3]. Additionally Logistic showed the highest values as 91.3 and 82.7 at the recall and F-measure (not described here). The recall means the ratio of the actual correct answers among the results of the computer has correctly



searched. In addition, Logistic showed the lowest value as 0.42 at RMSE. What RMSE has lower value means the difference between the actual and the expected value is small then it can be known that Logistic exhibits a relatively lower error rate than those of others. In the comparison of ROC analysis, Logistic has the highest value together with RandomF than others. It is known that the area under the ROC curve can be used in preference to overall accuracy for “single number” evaluation of machine learning algorithms [13]. When comparing between Kappa statistics values, it is shown to be near 0.2 in all the algorithms except MLP. As we know, Cohen's Kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories [3]. Therefore, even though in all the algorithms Kappa values showing near of 0.2, what the Kappa statistic is not 0 assures that the reliability of the system is good, and further it can be inferred that the precision is high as like the figure of bull's-eye [12]. The highest precision result of NB shown in Figure 3 appears again reflected in Figure 5. That is, in Figure 5, the specificity of NB is superior to others. However the sensitivity, which is often referred to as being related to the accuracy, is shown to be very low in NB. On the contrary to NB, Logistic and RandomF have best qualities in Sensitivity but worst in Specificity.

As described the above, the validity of data mining techniques were found to be different significantly depending on which algorithm or data are used. Also it was found to be variable depending on whether to use of any variable or how to use in the same material. In this paper, we concentrated upon the proposed algorithms, but it seemed to have a limitation that we only used the liver patient data set from India therefore could not assure if our findings were working rightly in another or various types of data set.

## 6. Conclusion

We evaluated, in this paper, the performance of classification algorithms considering the algorithmic characteristics to the liver patient data. For the evaluation of it, we performed the prediction test for comparing the results of our proposed algorithms with the originally used ones. Through our experiments and analysis, the effectiveness of data mining largely varied upon the algorithm used or the characteristics of data. Additionally more effective performance criteria should be accompanied with the choice of more appropriate algorithms. Lastly we conclude that it is important to choose effective algorithms and appreciate evaluation factors to the data by consideration of the algorithmic characteristics and the performance of prediction test respectively.

## References

- [1] B. Venkata Ramanal, M. S. P. Babu and N. B. Venkateswarlu, “A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis”, *International Journal of Database Management Systems (IJDMS)* vol. 3, no. 2, (2011), pp. 101-104.
- [2] [http://archive.ics.uci.edu/ml/datasets/ILPD+\(Indian+Liver+Patient+Dataset\)](http://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset)), (2013).
- [3] <http://en.wikipedia.org/>, (2013).
- [4] A. Martelli and U. Montanari, “Optimizing decision trees through Heuristically guided search”, *Commun. ACM*, vol. 21, no. 12, (1978), pp. 1025-1039.
- [5] B. H. Song, M. A. Jung and S. R. Lee, “A Design and Implementation Red Tide Prediction Monitoring System using Case Based Reasoning”, *Journal of Communication and Networks*, vol. 35, no. 12, (2010), pp. 1819-1826.
- [6] Y. G. Jung and H. Jin, “Experimental Comparisons of Neural Networks and Logistic Regression Models for Heart Disease Prediction”, *Information-An International Interdisciplinary Journal*, *International Informational Institute*, vol. 16, no. 2(B), (2013), pp. 1295-1300.
- [7] Y. B. Lee and S. H. Myaeng, “Text genre classification with genre-revealing and subject-revealing features”, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, (2002).

- [8] C. X. Ling and H. Zhang, "The representational power of discrete Bayesian network", *Journal of Machine Learning Research*, vol. 3, (2002), pp. 709-721.
- [9] K. P. Burnham and D. R. Anderson, "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach", Springer, (2002) July 12.
- [10] H. J. Choi, S. W. Choi and K. S. Han, "Prediction of DNA Binding Sites in Proteins using a Random Forest", *Journal of KIISE: Software and Applications* vol. 39, no. 7, (2012), pp. 515-522.
- [11] R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald and D. Scuse, "WEKA Manual for Version 3-6-0", The university of WAIKATO, (2008).
- [12] A. J. Viera and J. M. Garrett, "Understanding inter observer agreement: the kappa statistic", *Fam Med*, vol. 37, (2005), pp. 360-363.
- [13] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms", *Pattern Recognition*, vol. 30, no. 7, (1997), pp. 1145-1159.
- [14] G. Ming, "A comparison of Logistic Regression o Random Forests for Exploring Differences in Risk Factors Associated with Stage Atdiagnosis between Black and White Colon Cancer Patients", *Diss. University of Pittsburgh*, (2006).

## Authors



**Hoon Jin**, he is a research professor at College of Information & Communication Engineering in Sungkyunkwan University. He learned and studied Artificial Intelligence, in detail, data mining, bioinformatics, agent system, semantic web. He received the Ph.D. degree in Department of Computer Science of Kyonggi University, Suwon, Korea, in 2007. After graduation, He worked at Korea Research Institute of Bioscience and Biotechnology (KRIBB) in Daedeok Science Town of Daejeon, Korea during 3 years as a post-doctoral researcher. Then He worked as a senior researcher at Creative Design Institute aiming for studying the product service systems design in 2010. Recently he has worked as a research professor/senior researcher, at Yonsei Institute of Convergence Technology, Yonsei University for about 3 years. His main research interests are in fields of data mining, bioinformatics, ontology, semantic web technologies and now actively researches about biomedical text mining.



**Seungcheon Kim**, He has received the B.S., M.S. and Ph.D. degrees in Electronic Engineering Department of Yonsei University, Seoul, Korea, in 1994, 1996 and 1999, respectively. He is currently with the Department of Information and Communication Engineering, Hansung University, Seoul, Korea, where he is responsible for teaching and research in wireless data communication networks, and ubiquitous sensor networks. He has worked as a post doctoral research fellow in the School of Electrical and Information Engineering in the University of Sydney, Australia, from 2000 to 2001, where he conducted research about 4G Mobile Wireless Communications. He's also worked as a senior research engineer in the Home Network Group of Digital TV Laboratory and the Digital Tech. Group of DA Laboratory, LG Electronics Inc., from 2001 to 2003, where he designed the Home Network Protocol and developed several Home Networking Devices. He has served as a director of Industrial cooperation research center in Hansung

University. He was a visiting scholar in the department of computer science in the University of Oregon, United States, from 2009 to 2010. His research interests include the traffic managements in Wireless and mobile communication networks, architectures of 4G Wireless Networks and the design of Home Networking Protocol and Ubiquitous Network Architecture.



**Jinhong Kim**, he is Associate Professor of Department of Computer Engineering at the Hansung University, Seoul, and Republic of Korea. He respectively received his Ph.D. degrees in Computer Engineering from Sungkyunkwan University (SKKU), Korea, in 2006. Dr. Kim was a post doctorate research associate of the Software Engineering at the SKKU, as well as of the Information System Research Center at the University of New South Wales in Australia. Moreover, He acted as an Engineer expert on behalf of Vodafone Company. He has served on a Co-chair of the IEIE Computer Society. He also served or currently serving as a reviewer and Technical Program Committee for many important Journals, Conferences, Symposiums, Workshop in Computer Communications Networks area. His research interests include smart vehicular network, smart platform, software engineering, wireless sensor networks, scalable reliable communication protocols, mobile computing, network security protocols, proxy caching systems, and formal verification of communication protocols.