

Rule Discovery for Diabetes Mellitus Diagnosis using Ant-Miner Algorithm

SARITCHAI PREDAWAN¹

KORRAKOT SINPRAJAKPOL² TERAPONG AYAMUANG³

Sirindhorn College of Public Health Chonburi

29 moo. 4, Bansuan, Muang, Chonburi Thailand, 20000

¹saritchai@scphc.ac.th ²luktao@gmail.com ³teraponga@gmail.com

Abstract: - Diabetes mellitus currently affects over 425 million people worldwide. According to the WHO report, by 2045[3] this number is expected to rise to over 629 million. The disease has been named the 2nd of NCDs (Non-Communicable diseases) in Thailand. In diagnosis of Diabetes mellitus are done mostly by expertise and experienced doctors, but still there are cases of wrong diagnosis. Patient have to undergo various test which are very costly and sometimes all of them are not required so in this way it will hugely increase the bill of a patient unnecessarily. This paper presents diabetes mellitus diagnosis system by analyzing the patterns via Pima Indian Diabetes Dataset (PIDD). The system is composed of main process, Pima Indian Diabetes Dataset are cleaned and transformation. Normal distributions are employed by Z-transform function. In rule discovery for diagnosis, we used the Ant-Miner classifier to classify Diabetes by assuming that the feature is features Diagnosis. This experiment, Ant-Miner algorithm is adapted, with a small change to increase the accuracy rate. The result of this experiment is more than 86% accuracy rate and shows that the constructed data mining model could assist health care providers to make better clinical decisions in identifying diabetic patients.

Key-Words: - Pima Indian Diabetes Dataset (PIDD), Normalization, Diabetes Mellitus, Ant Colony Optimization (ACO)

1 Introduction

Diabetes Mellitus (DM) or Diabetes is a chronic disease which is characterized by hyperglycemia that results from deficiency in insulin secretion, insulin action, or both, is a group of metabolic diseases. DM has two types. The pathogenesis of type 1 diabetes mellitus (T1DM) is that the pancreas secretes damaged β -cells, preventing it from lowering blood glucose level in time. Insulin resistance and insulin secretion deficiency are the pathogenesis of type 2 diabetes mellitus (T2DM) [1].

Result of "Prevalence of Diabetes and Relationship with Socioeconomic Status in The Thai Population: National Health Examination Survey, 2004-2014" found epidemic Diabetes or Diabetes Mellitus (DM) continues to grow in the Thai population, especially in individuals with lower education. New cases and strengthen glycemic control should be scaled up. Thailand is facing financial challenge due to the prevalence of DM. The cost includes medications, complication diseases, visits, and lab tests, which also varies based on the patient's country. Prevention,

monitoring and controlling are the most effective actions to face such a health care challenge [2].

IDF Diabetes Atlas shown the number of diabetic worldwide was 425 million. The population of DM was continuously increasing occurrence, it predicts that the number will approach to 629 million in 2045[3]. It is vital for us to focus on a high-risk group of people with DM to lower the morbidity. According to the World Health Organization (WHO) standard(2016), the definitions of groups with high risk of DM are as follows:[4]

- Age ≥ 45 and rare exercising
- BMI ≥ 24 kg/m²
- Impaired glucose tolerance (IGT) of impaired fasting glucose (IFG)
- Family history of DM
- Lower high-density lipoprotein cholesterol or hypertriglyceridemia (HTG)
- Hypertension or cardiovascular and cerebrovascular disease
- Gestation female whose age ≥ 30

For medical research due to the longevity of the DM and the huge cost on the health care provides. Early detection of diabetes ultimately reduces cost

on health care providers for treating diabetic patient [5], [6], [7]. Early Detection of DM, researcher can take advantage of the high-risk group of DM data to convert raw data into meaningful information and extract hidden knowledge. We need to utilize advance information technology. Therefore, data mining technology is an appropriate study field for us. Data mining, also known as Knowledge Discovery in Database (KDD), is defined as the computational process of discovering patterns in large datasets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems [8],[9].

In this paper, efforts were Ant-Miner algorithm on data mining methods used in diabetes mellitus classification. Therefore, believe it necessary to establish a model that can classify patients into either suspected patients of confirmed patients in high-risk DM.

The remaining paper consists of the following section; firstly, the related works and data source are given in section 2, the section 3 presents data preprocessing and ant colony optimization, experimental and result are given in section 4. Finally, concludes the paper with some directions for future work.

2 Related works and Data Source

There are various health care studies providing useful informative knowledge of algorithm approaches in data mining and classification rule discovery purposes. In this section, a few important works that are closely related to the proposed issue are presented. With the evolution of information technology, data mining provides a valuable asset in diabetes research, which leads to improve health care delivery, increase support to decision-making and enhance disease management [10].

On several researches described the Pima Indian Diabetes Dataset (PIDD) from the University of California, Irvine [11] using various data mining and soft computing methods. This benchmark data set is mostly used to decision support systems in diabetes diagnosis and to compare the prediction models. K.Kayaer and T.Yildirim [12] proposed various Artificial Neural Network (ANN) and stated that General Regression Neural Network (GRNN) was good to classify diabetes data set compared with Multilayer Perceptron (MLP) and Radius Basis Function (RBF) with accuracy of 80.21%. Patil [13] proposed a Hybrid Prediction Model (HPM), which used a K-means clustering algorithm aimed at validating a chosen class label of given data and

used the C4.5 algorithm aimed at building the final classifier model, with 92.38% classification accuracy. Ahmad [14] compared the prediction accuracy of MLP in neural networks against the ID3 and J48 algorithms. The results showed that a pruned J48 tree performed with higher accuracy, which was 89.3% compared to 81.9%. All methods presented above used the same Pima Indian diabetes data set as the experimental data source. The Waikato Environment for Knowledge Analysis (WEKA) toolkit was the major tool which most researchers use. From the literature review shows data mining algorithms have been used to predict diabetes using public data.

Ant Colony Optimization (ACO) is very effectively and successively applied for the discovery of classification rule mining. There are a vast variety of application of Ant Colony Optimization in the field of data mining. Ant Colony Algorithms for speech recognition is proposed in S.Predawan [15]. Sonal P. Rami and Mahesh H. Panchil [16] presented some dialects and Ant-Miner by using public data for the observation of impact of number of ants on the accuracy rate. N.N Das and Anjali Saini [17] proposed algorithms for association rule mining and basics terms of Ant Colony Optimization algorithms are discussed. S. vanaja and K. Rameshkumar [18] analyzed performance of classification algorithms on various medical data sets. On the basis of important and effectiveness of knowledge discovery from huge data reservoirs, discussions were provided by the A.A Freitas [19], The study of Ant Colony Algorithms for data classification is given in [20].

This literature survey contributed to research society by providing larger number of Rule Miners exploiting Ant Colony Optimization particularly and updated related work continued for the discovery of classification rules.

2.1 Data Source

In this study, the general framework of proposed model is shown in Figure. 1. The proposed method consists of several steps, which are collection of data set, attribute selection, normalized values, data mining algorithms and achievement of learning models.

Dataset Description

The Pima Indian Diabetes Datasets were selected from a larger data set held by the National Institutes of Diabetes and digestive and Kidney Disease.

All patients in PIDD are women at least 21 years old and living near Phoenix, Arizona, USA. The

data sets consist of 768 diabetic and non-diabetic instances. Each instance is consisted of 8 attributes. The descriptions of the attributes in the dataset are shown in Table 1 and Table 2 shows some raw data from the PIDD.

Table 1. All attributes Description.

Attribute	Relabeled Attributes	Min/Max Values
1. Number of times pregnant	N_Preg	0/17
2. Plasma glucose concentration at 2 hr. in an oral glucose tolerance test	P_Gluc	0/199
3. Diastolic blood pressure (mm Hg)	B_Pres	0/122
4. Triceps skin fold thickness (mm)	T_Skin	0/99
5. 2-Hour serum insulin (mu U/ml)	H_Insu	0/846
6. Body Mass Index (kg/m ²)	BMI	0/67.1
7. Diabetes Pedigree Function	D_Pedi	0.078/2.42
8. Age in Years	AGE	21/81

Table 2 Raw data of PIDD.

N_preg	P_Gluc	B_Pres	T_Skin	H_Insu	BMI	D_Pedi	Age
1	122	90	51	220	49.7	0.325	31
1	163	72	0	0	39	1.222	33
1	151	60	0	0	26.1	0.179	22
0	125	96	0	0	22.5	0.262	21
1	81	72	18	40	26.6	0.283	24
2	85	65	0	0	39.6	0.93	27

Extensive methods by many researchers have also been done on Pima Indian Diabetes Dataset, and the results obtained are presented in Table 3.

Table 3. The values of accuracy of classification on PIDD.

Algorithm	Accuracy (%)	Reference
RBF	68.23	Kayaer, Yidirim [12]
MLP	77.08	Kayaer, Yidirim [12]
GRNN	80.21	Kayaer, Yidirim [12]
k-NN	71.90	Ster. Dobnikar [21]
CART	72.80	Kayaer, Yidirim [12]
MLP	75.20	Kayaer, Yidirim [12]
LVQ	75.80	Kayaer, Yidirim [12]
Navie Bayes	74.50	Friedman [22]
Semi-naïve Bayes	76.00	Friedman [22]
C4.5	76.00	Friedman [22]

Legend: PIDD: Pima Indian Diabetes Dataset. RBF: radial basis function. MLP: Multi-layer perceptron. GRNN: general regression neural network. k-NN: k-nearest-neighbor. CART: classification and regression trees. LVQ: learning vector quantization. C4.5: sample class 4.5 algorithm.

3 Data preprocessing and Ant Colony Optimization

Raw data is highly susceptible to noise, missing values, and inconsistency. The quality of data affects the data mining results. In order to help improve the quality of the data and consequently, of the mining results raw data is pre-processed so as to improve the efficiency and ease of the mining process. Data preprocessing is one of the most critical step in a data mining process which deals with preparation and transformation of the initial dataset.

3.1 Data Cleaning

Incomplete data can occur for a number of reasons. Attributes of interest may not always be available. Relevant data may not be recorded due to a misunderstanding, or because of equipment malfunctions. Data that were inconsistent with other recorded data may have been deleted. Furthermore, the recording of the history or modifications to the data may have been deleted. Missing data, particularly for tuples with missing values for some attributes, may need to be inferred. Data can be noisy, having incorrect attribute values, owing to the following. Incorrect data may also result from inconsistencies in naming conventions or data codes used. Duplicate tuples also require data cleaning. Data cleaning routines work to "Clean" the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Dirty data can cause confusion for the mining procedure.

3.2 Data Transformation

The data are consolidated by Z-transform function into forms appropriate for mining. Data transformation can involve the following:

- Normalization, where the attribute data are scaled so as to fall within a small specified range, such as -1.0 to 1.0, of 0 to 1.0.

- Aggregation, where summary or aggregation operations are applied to the data. For example, the daily sales data may be aggregates so as to compute monthly and annual total amounts. This step is

typically used in constructing a data cube for analysis of the data at multiple granularities.

- Generalization of the data, where low level of “primitive” (raw) data are replaced by higher level concepts through the use of concept hierarchies. For example, categorical attributes, like street, can be generalized to higher level concepts, like city or county. Similarly, values for numeric attributes, like age, may be mapped to higher level concepts, like young, middle-aged, and senior.

3.3 Ant Colony Optimization

Ant Colony Optimization (ACO) is a branch of a newly developed form of artificial intelligence called swarm intelligence. Swarm intelligence is a field which studies “the emergent collective intelligence of groups of simple agents” [23]. In groups of insects, which live in colonies, such as ants and bees, an individual can only do simple tasks on its own, while the colony's cooperative work is the main reason determining the intelligent behavior it shows. Most real ants are blind. However, each ant while it is walking, deposits a chemical substance on the ground called pheromone [24]. Pheromone encourages the following ants to stay close to previous moves. The pheromone evaporates over time to allow search exploration. In a number of experiments presented in [24], Dorigo and Maniezzo illustrate the complex behavior of ant colonies.

For example, a set of ants built a path to some food. An obstacle with two ends was then placed in their way such that one end of the obstacle was more distant than the other. In the beginning, equal numbers of ants spread around the two ends of the obstacle. Since all ants have almost the same speed, the ants going around the nearer end of the obstacle return before the ants going around the farther end (differential path effect). With time, the amount of pheromone the ants deposit increases more rapidly on the shorter path, and so more ants prefer this path. This positive effect is called autocatalysis. The difference between the two paths is called the preferential path effect; it is the result of the differential deposition of pheromone between the two sides of the obstacle, since the ants following the shorter path will make more visits to the source than those following the longer path. Because of pheromone evaporation, pheromone on the longer path vanishes with time. The goal of Ant-Miner is to extract classification rules from data [25]. The algorithm is presented in Figure 1.

3.3.1 Pheromone Initialization

All cells in the pheromone table are initialized equally to the following value (1):

$$\tau_{ij(0)} = \frac{1}{\sum_{i=1}^a b_i} \quad (1)$$

where:

- a is the total number of attributes;
- b_i is the number of values in the domain of attributes A_i.

3.3.2 Rule Construction

Each rule in Ant-Miner contains a condition part as the antecedent and a predicted class. The condition part is a conjunction of attribute-operator-value tuples. The operator used in all experiments is “=”, just as in Ant-Miner, all attributes are assumed to be categorical. Let us assume a rule condition such as term $ij \approx A_i = V_{ij}$, where A_i is the ith attribute and V_{ij} is the jth value in the domain of A_i. The probability, that this condition is added to the current partial rule that the ant is constructing, is given by the following Equation (2):

$$P_{ij}(t) = \frac{\tau_{ij}(t) \cdot \eta_{ij}}{\sum_i^a \sum_j^{b_i} \tau_{ij}(t) \cdot \eta_{ij}} ; \forall i \in I \quad (2)$$

Where:

- η_{ij} is a problem-dependent heuristic value for term ij
- $\tau_{ij}(t)$ is the amount of pheromone currently available(at time t) on the connection between attribute i and value j
 - a is the total number of attributes
 - b_i is the total number of values in the domain of attributes A_i
 - I is the set of attributes that are not yet used by the ant

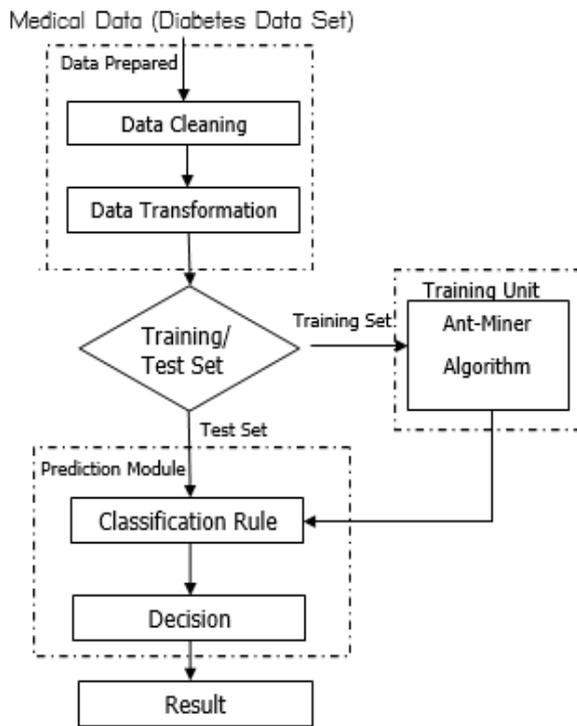


Figure 1. Proposed Diagram

3.3.3 Heuristic Value

In traditional ACO, a heuristic value is usually used in conjunction with the pheromone value to decide on the transitions to be made. In Ant-Miner, the heuristic value is taken to be an information theoretic measure for the quality of the term to be added to the rule. The quality here is measured in terms of the entropy for preferring this term to the others, and is given by the following equations (3), (4):

$$\eta_{ij} = \frac{\log_2(k) - InfoT_{ij}}{\sum_i^a \sum_j^b \log_2(k) - InfoT_{ij}} \quad (3)$$

$$InfoT_{ij} = -\sum_{w=1}^k \left[\frac{freqT_{ij}^w}{|T_{ij}|} \right] * \log_2 \left[\frac{freqT_{ij}^w}{|T_{ij}|} \right] \quad (4)$$

where:

- k is the number of classes
- |Tij| is the total number of cases in partition Tij (partition containing the cases where attribute Ai has value Vij)
- freq T w ij is the number of cases in partition Tij with class w

The higher the value of info Tij, the less likely that the ant will choose term ij to add to its partial rule.

3.3.4 Rule Pruning

Immediately after the ant completes the construction of a rule, rule pruning is undertaken to increase the comprehensibility and accuracy of the rule. After the pruning step, the rule may be assigned a different predicted class based on the majority class in the cases covered by the rule antecedent. The rule pruning procedure iteratively removes the term whose removal will cause a maximum increase in the quality of the rule. The quality of a rule is measured using the following equation(5):

$$Q = \left(\frac{TruePos}{TruePos + FalseNeg} \right) \times \left(\frac{TrueNeg}{FalsePos + TrueNeg} \right) \quad (5)$$

Where:

- TruePos is the number of cases covered by the rule and having the same class as that predicted by the rule
- FalsePos is the number of cases covered by the rule and having a different class from that predicted by the rule
- FalseNeg is the number of cases that are not covered by the rule, while having the class predicted by the rule
- TrueNeg is the number of cases that are not covered by the rule which have a different class from the class predicted by the rule.

3.3.5. Pheromone Update Rule

After each ant completes the construction of its rule, pheromone updating is carried out as follows (6) :

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t) \cdot Q, \forall term_{ij} \in \text{the rule} \quad (6)$$

To simulate the phenomenon of pheromone evaporation in real ant colony systems, the amount of pheromone associated with each term ij which does not occur in the constructed rule must be decreased, The reduction of pheromone of an unused term is performed by dividing the value of each τij by the summation of all τij.

4 Experimental and Result

Data gathering methods are often loosely controlled, resulting in out of range values, impossible data combinations (e.g., AGE: 18, Pregnant: 4), missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis.

Table 4. Descriptive statistics of PIDD

Attributes	Mean	Std.Deviation	Min/Max
N_Preg	3.8	3.4	0/17
P_Gluc	20.91	32.0	0/199
B_Pres	69.1	19.4	0/122
T_Skin	20.5	16.0	0/99
H_Insu	79.8	115.2	0/846
BMI	32.0	7.9	0/67.1
D_Pedi	0.5	0.3	0.078/2.42
AGE	33.2	11.8	21/81

In this study, data preprocessing plays an important value in our proposed model. Each attribute's diabetes dataset were analyzed and its correlation to diabetes mellitus. In N_Preg value of 0, which transformed into a nominal value with "no". Transformation steps of the PIDD include replacing missing value and normalization of values. There are some missing and high value ranges between the attributes in the dataset due to deregulation. Several inaccurate researches' results were caused by these values.

A normalization process in performed on the data to fix this problem and to get the best results. In data normalization, the data are transformed or consolidated into forms appropriate for data mining. The attribute data are scaled so as to fall within a small specified range, such as 0 to 1.0. The statistics of the PIDD are presented in Table 4. The normalized statistic values of PIDD is shown in Table 5.

Table 5. Normalized statistical values of PIDD

Attributes	Mean	Std.Deviation	Min/Max
N_Preg	3.80	3.40	0/17
P_Gluc	12.09	3.20	0/199
B_Pres	6.91	1.94	0/122
T_Skin	2.05	1.60	0/99
H_Insu	0.798	1.152	0/846
BMI	3.20	0.79	0/67.1
D_Pedi	5	3	0.078/2.42
AGE	3.32	1.18	21/81

The platform used for conducting following experiments is a personal computer (PC). The algorithm is implemented by Java, the experimental parameters are confirmed by large number of experiments. Our system has the following four user-defined parameters:

- Number of ants (λ): This is also the maximum number of complete candidate rules constructed and pruned during an iteration of the WHILE loop of Ant-Miner's Algorithm, since each ant is associated with a single rule. In each iteration,

the best candidate rule found is considered a discovered rule. The larger, the more candidate rules are evaluated per iteration, but the slower the system is.

- Minimum number of cases per rule (μ): Each rule must cover at least cases to enforce at least a certain degree of generality in the discovered rules. This helps to avoid an over fitting of the training data.

- Maximum number of uncovered cases in the training set (ν): The process of rule discovery is iteratively performed until the number of training cases that are not covered by any discovered rule is smaller than this threshold.

- Number of rules used to test convergence of the ants (ρ): If the current ant has constructed a rule that is exactly the same as the rule constructed by the previous ants, then the system concludes that the ants have converged to a single rule (path). The current iteration of the WHILE loop of Ant-Miner's Algorithm is therefore stopped and another iteration is started.

We assess the performance of the improved algorithm, and compare its test results to Ant-Miner under the same experimental environment. All the experiments reported in this paper these parameters were set as follows:

No_of_ants = 50,100 and 250 respectively.

Min_cases_per_rule = 5.

Max_uncover cases = 10.

No_Rules_Convergence = 10.

Evaporation Factor (ρ) = 0.95.

α and β = 1.

Table 6 shows the main Characteristic of the datasets, which were used to evaluate in our system.

Table 6. DataSet Characteristics

DataSet	Number of examples	Number of attributes	Number of classes
PIDD	768	8	2

For all datasets, a 10-fold cross validation was used. In this procedure, all cases are used only once as testing and training. The final accuracy rate is simply the average of the accuracy rate of the iterations. All the data partitions are randomly generated considering all available cases.

Table 7. Results of Testing with the Ant-Miner Algorithm

No_of_Ant	Accuracy Rate	Rules Number	Conditions Number
50	84.61% \pm 1.15	9.7 \pm 0.21	14.7 \pm 0.65
100	83.83% \pm 1.20	9.5 \pm 0.22	14.2 \pm 0.59
250	91.48% \pm 2.24	9.5 \pm 0.17	14.1 \pm 0.72

Table 7 summarizes the results of testing obtained by the proposed Ant-Miner algorithm in our datasets. The table shows the accuracy rate, the number of rules found and the number of terms (the shown values are the average values of the cross-validation procedure followed by the corresponding standard deviation). The result shows that the Ant-Miner parameter is able to produce minimum rules for diagnosis with more than 86% accuracy rate. The average number of rules used is 9.

5 Conclusions

We have described an Ant Colony System called Ant-Miner for the discovery of classification rules in databases [25]. We have also shown results indication that Ant-Miner had a good classification performance on the datasets used in our experiments. These results also show that the proposed algorithm is able to achieve both good predictive accuracy and a reduced number of rules at the same time. This facilitates the practical use the system, since it usually generates comprehensible rules. The main drawback is still its computational cost, especially when the search space (number of predicting attributes) is too large. However, to develop the understanding of parameters and effects of each parameter of every system needs a very detailed experimentation. The result of this paper show that the constructed data mining model could assist health care providers to make better clinical decisions in identifying diabetic patients.

As future research direction, it would be interesting to investigate the performance of different feature selection methods. Also it would be interesting to compare the proposed classifier with other classifiers and evaluate its performance on different data sets. Finally, it would be promising to automate the procedure by embedding the feature selection algorithm to the ACO classification algorithm.

References :

- [1] A. Nanditha, R. Ma, A. Ramachandran, C. Snehalatha, J. Chan, K. Chia, J. Shaw and P. Zimmet, "Diabetes in Asia and the Pacific: Implications for the Global Epidemic.," *Diabetes Care*, pp. 472-485, 2016.
- [2] W. Akeplakorn, S. Chariyalertsak, P. Kessomboon, S. Assanangkornchai, S. Taneepanichskul and P. Putwatana, "Prevalence of Diabetes and Relationship with Socioeconomic Status in Thai Population: National Health Examination Survey, 2004-2014," *J Diabetes Res.*, p. PMC5852889, 2018.
- [3] I. D. Federation, "IDF Diabetes Atlas 8th Edition," International Diabetes Federation, UK, 2017.
- [4] World Health Organization, "Global Report on Diabetes," World Health Organization, France, 2016.
- [5] R. Li, P. Zhang, L. Barker, F. Chowdhury and X. Zhang, "Cost-effectiveness of interventions to prevent and control diabetes mellitus: a systematic review.," *Diabetes Care*, pp. 1872-1894, 2010.
- [6] J. Lin and P. Haug, "Data preparation framework for preprocessing clinical data in data mining.," *AMIA Annual Symposium Proceeding*, pp. 489-493, 2006.
- [7] M. Luboschik, M. Rohlig, G. Kundt, O. Stachs, S. Peschel, A. Zhivov, R. F. Guthoff, K. Winter and H. Schumann, "Supporting an Early Detection of Diabetic Neuropathy by Visual Analytics," *EuroVis Workshop on Visual Analytics*, pp. 49-53, 2014.
- [8] S. Nuwangi, C. Oruthotaarachchi, J. Tilakaratna and H. Caldera, "Utilization of Data Mining Techniques in Knowledge Extraction for Diminution of Diabetes," in *2010 Second Vaagdevi International Conference on Information Technology for Real World Problems.*, Warangal, India, 2010.
- [9] U. M. Fayyad, P.-s. Gregory, S. Fadhraic and U. Ramasamy, *Advances in knowledge discovery and data mining.*, CA, USA: American Association of Artificial Intelligence Menlo Park, 1996.
- [10] H. Jiawei, K. Micheline and P. Jian, *Concepts and Techniques*, 3rd Ed., USA: Morgan Kaufmann Publisher., 2012.

- [11] M. Rodrigo and Silvio do La go Pereira, "Evolutionary Learning of Concepts," *Journal of Computer and Communications.*, pp. 76-86, 2014.
- [12] K. Kamer and Y. Tulay, "Medical diagnosis on Pima Indian diabetes using general regression neural networks," in *Proceeding of The International Conference on Artificial Neural Networks and Neural Information Processing*, Iran, 2003.
- [13] B. Patil, "Hybrid prediction model for Type-2 diabetic patients.," in *Expert System Applied*, India, 2010.
- [14] A. Aliza, M. Aida, D. Z. Eliza, M. Norhayati and Y. Y. Nur, "Comparison between Neural Networks against Decision Tree in Improving Prediction Accuracy for Diabetes Mellitus.," in *ICDIP 2011: Digital Information Processing and Communications.*, Ostrava, Czech Republic, 2011.
- [15] S. Predawan, P. Jiyapanichkul, C. Kimpan and C. Wutiwiatchai, "A Method for Thai Isolated Word Recognition using Ant Colony Algorithm.," in *5th WSEAS Int. Conf on Computational Intelligence MAN-Machine Systems and Cybernetics.*, Venice, Italy, 2006.
- [16] P. R. Sonal and H. P. Mahesh, "Comparative Analysis of Variations of Ant-Miner by Varying Input Parameters.," *International Journal of Computer Applications*, pp. 25-29, 2012.
- [17] N. Das and S. Anjali, "A Survey on Different Algorithm of Association Rule Mining and Ant Colony Optimization.," *International Journal of Information Science and Computing.*, pp. 43-48, 2014.
- [18] S. Vanaja and K. Pameshkumar, "Performance Analysis of Classification Algorithms on Medical Diagnoses-a Survey," *Journal of Computer Science.*, pp. 30-52, 2014.
- [19] A. A. Freitas, "Are We Really Discovering "Interesting" Knowledge From Data?," *Expert Update In the BCS-SGAI Magazine.*, pp. 41-47, 2006.
- [20] A. A. Freitas, "A Review of evolutionary Algorithms for Data Mining.," in *Soft Computing for Knowledge Discovery and Data Mining.*, Boston, MA, Springer, 2008, pp. 79-111.
- [21] S. Branko and D. Andrej, "Neural Networks in medical diagnosis: Comparison with other methods," in *International Conference on Engineering Application with Neural Networks.*, London, 1990.
- [22] N. Friedman, D. Geiger and M. Glodszmidt, "Bayesian Network Classifiers.," in *Machine Learning*, Netherlands, Kluwer Academic Publishers, 1997, pp. 131-163.
- [23] R. S. Parpinelli, H. S. Lopes and A. Freitas, "An Ant Colony Algorithm for Classification Rule Discovery.," in *Data mining: A Heuristic Approach.*, New South Wales, Australia, Idea Group Publishing, 2002, pp. 181-208.
- [24] M. Dorigo, G. Di Caro and L. Gambardella, "Ant algorithms for discrete optimization.," *Artificial Life.*, pp. 137-172, 1999.
- [25] M. Dorigo, V. Maniezzo and A. Colorni, "Ant system: optimization by a colony of cooperation agents.," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, pp. 29-41, 1996.