Fourth International Workshop on Computational Intelligence & Applications
IEEE SMC Hiroshima Chapter, Hiroshima University, Japan, December 10 & 11, 2008

Classification results of coronary heart disease database by using the clonal selection method with receptor editing

Yasuaki Sakamoto[1)], Takumi Ichimura[2)]*, Akira Hara[2)], and Tetsuyuki Takahama[2)]
1) Department of Intelligent Systems, Faculty of Information Sciences, Hiroshima City University
2) Graduate School of Information Sciences, Hiroshima City University
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima 731-3194 Japan
E-mail : ysakamoto@chi.its.hiroshima-cu.ac.jp, {ichimura, ahara, takahama}@hiroshima-cu.ac.jp
*Corresponding author

***Abstract*—The clonal selection principle is used to explain the basic features of an adaptive immune response to a antigenic stimulus. It established the idea that only those cells that recognize the antigens are selected to proliferate and differentiate. This paper explains a computational implementation of the clonal selection principle that explicitly takes into account the affinity maturation of the immune response. The clonal selection algorithm by incorporating receptor editing method, RECSA, has been proposed by Gao. This paper tries to classify the medical database of Coronary Heart Disease databases and reports the computational results for 4 kinds of training datasets.**

I. INTRODUCTION

In a few decades, the area of artificial immune system (AIS) has been an ever-increasing interested in not only theoretical works but applications in pattern recognition, network security, and optimizations[1],[2],[3],[4],[5],[6],[7]. AIS uses ideas gleaned from immunology in order to develop adaptive systems capable of performing a wide range of tasks in various research areas. Especially, Gao focuses the Clonal Selection Algorithm(CSA) which is known to one of the famous immune algorithms and proposed the effective method for solving Traveling Salesman Problem(TSP)[8] which is known to be NP-complete.

The biological basis of the Clonal Selection Theory was proposed by Burent[13],[14] in 1959. The theory interprets the response of lymphocytes in the face of antigenic stimulus. Only the immune cells with high affinity are selected to proliferate, while those low affinity cells must be efficiently deleted or become anergic. The hypermutation is allowed to improve the affinity of the selected cells to the selective antigens. Receptor Editing as a mechanism of immune cell tolerance is reported[9],[10],[11].

Gao indicated the complementary roles of somatic hypermutation (HM) and receptor editing (RE) and presented a novel clonal selection algorithm called RECSA model by incorporating the Receptor Editing method[8]. In [8], they discussed the relationships between HM and RE through utilizing them to solve the TSPs. Because a valid tour in TSP is represented by a permutation of N cities, the number of states to feasible tours is $(N-1)!$. [1] solves the problems for finding an optimal set in the search space consisted of the set of $(N-1)!$ valid tours in TSPs.

A medical database named Coronary Heart Disease Database (CHD_DB) has been prepared in the data mining contest. The database makes it possible to assess the effectiveness of data mining method in medical data. The CHD_DB is based on actual measurements of the Framingham Heart Study - one of the most famous prospective studies of cardiovascular disease. It includes more than 10,000 records related to the development of coronary heart disease (CHD). The datasets have been proved enough valid by statistical analyses[12].

This paper challenges to classify the CHD_DB by using RECSA model. The classification of medical database differs from the TSP, because medical information such as results of biochemical tests and chief complaint is often ambiguous. Therefore, we cannot clearly distinguish the difference between normal and pathological values. Biochemical test values cannot be precisely evaluated by using crisp sets. In this paper, we consider that the database has some relation between inputs and output signals. That is, an output is summed up all the inputs modified by their respective weights and is compared with the corresponding teach signal. We report the computational classification results of the databases.

The remainder of this paper is organized as follows. In the next serction, the clonal selection theory will be explained briefly. Section 3 will explain RECSA model proposed by [8]. The CHD_DB will be described in Section 4. Experimental results for classification of the CHD_DB will be reported in Section 5. In Section 6, we give some discussions to conclude this paper.

II. THE CLONAL SELECTION THEORY

Burnet proposed the clonal selection theory in order to explain the essential features of adaptive immune response [13][14]. The basic idea of this theory interprets the response of lymphocytes in the face of an antigenic stimulus. Fig.1 shows the overview of clonal selection principal.

Any molecule that can be recognized by the adaptive immune system is known as antigens (*Ag*s). Some subpopula-

tions of its bone-marrow-derived cells responds by producing antibody (*Ab*). *Ab*s are molecules attached primarily to the surface of B cells. The aim of B cell is to recognize and bind to *Ag*s. Each B cell (B lymphocytes) secretes a single type of *Ab*. By binding to these *Ab*s and with a second signal from T-helper cell, the *Ag* stimulates the B cell to proliferate and mature into terminal *Ab* secreting cells called plasma cells. Proliferation of the B cell is a mitotic process whereby the cells divide themselves, creating a set of clones identical to the parent cell. The proliferation rate is directly proportional to the affinity level. That is, the higher affinity level of B cells, the more of them will be readily selected for cloning and cloned in larger numbers.

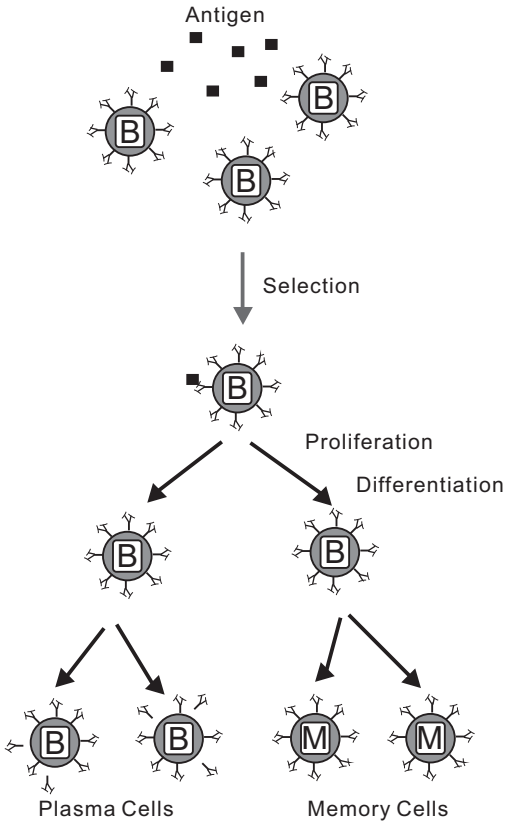

Fig. 1. An overview of the clonal selection principle

In addition to proliferating and differentiating into plasma cells, B cells can differentiate into long lived B memory cells. Memory cells circulate through the blood, lymph, and tissues. By exposing to a second antigenic stimulus, they commence to differentiate into plasma cells capable of producing high affinity *Ab*s, which are preselected for the specific *Ag* that stimulated the primary response.

Gao has been proposed the clonal selection method by incorporating receptor editing which can diversify the repertoire of antigen activated B cells during asexual reproduction[8]. The proposed method has two mechanisms of somatic hypermutation and receptor editing [15][16][17].

First, the basic idea of somatic hypermutation is explained. A rapid accumulation of mutations is necessary for a fast maturation of immune response. A large proportion of the cloned population becomes dysfunctional or develops into harmful anti-self cells after mutation. However, an effective change enables the offspring cell to bind better with antigen, then affinity is improved. When a mutated cloned cell with higher affinity is found, it will be activated to undergo proliferation. The mutation in the cloned cells happens with inverse proportional to the antigen affinity. This process of constant selection and mutation of only the B cells with antibodies is affinity maturation. The immune system is capable of evolving antibodies to recognize and bind with not only known antigens but unknown ones by affinity maturation. However, those cells with low affinity may be further mutated.

Second, receptor editing, the main role in shaping the lymphocyte repertoire, is explained. Both B cell and T cell that carry antigen receptors are able to change specificity through subsequent receptor gene rearrangement. The antigen molecule is composed of two chains, each resulting from the somatic rearrangement of various genetic segments as shown in Fig. 2 and Fig. 3. The B cell expresses at their surface a receptor allowing them to specifically recognize antigens. These figures show the rearrangement at the stage of the pro-B cells. In the primal rearrangement as shown in Fig.2, the heavy chain locus is rearranged to produce a *VDJ* segment. It encodes numerous *V*(variable), *D*(Density), *J*(Junction) segment. When one *D* segment and one *J* segment join together, a *V* segment is then joined to the assembled *DJ*. This pattern forms a unique combination of *VDJ*. In the secondary rearrangement as shown in Fig.3, the process governing B cell tolerance is the light chain. In case locus lack *D* region, only *V* and *J* segment will be assembled. After an initial *VJ* rearrangement, upstream *V* segment can be further rearranged to downstream J segment, deleting or displacing the previously rearranged *VJ* segment. Furthermore, numerous *V* regions are in reverse orientation on chromosome, these *V* are rearranged by inversion rather than by deletion of the intervening sequences.

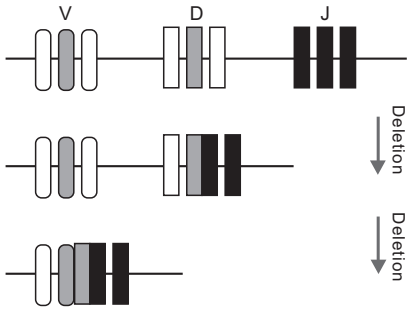

Fig. 2. The arrangement of *VDJ* (a)heavy chain

In addition to somatic hypermutation and receptor editing, it should be note that some of stimulated lymphocytes is replaced per cell generation by newcomer cells from the bone marrow.

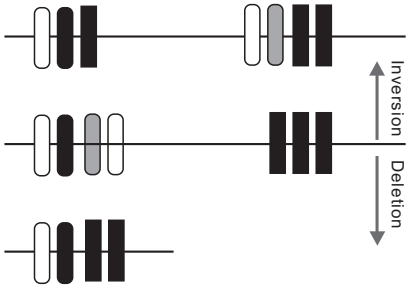

Fig. 3. The arrangement of *VDJ* (b)light chain

III. RECSA MODEL[8]

The shape-space model aims at quantitatively describing the interactions among *Ag*s and *Ab*s (*Ag-Ab*) [18][19]. The set of features that characterize a molecule is called its generalized shape. The *Ag-Ab* codification determines their spatial representation and a distance measure is used to calculate the degree of interaction between these molecules. Mathematically, the generalized shape of a molecule (m) can be represented by a set of L attributes directly associated with coordinate axes. These axes are represented by $m = \langle m_L, \cdots, m_2, m_1 \rangle$ which can be regarded as a point in an L-dimensional real-valued shape space $S, m \in S^L \subset \Re^L$. So we can express the immune cell's receptor gene sequence as $R = (r_1, r_2, \cdots, r_L)$.

The Gao's model[8] as shown in Fig. 4 can be described as follows.

Step1) Create an initial pool of m antibodies as candidate solutions$(Ab_1, Ab_2, \cdots, Ab_m)$.

Step2) Compute the affinity of all antibodies: $(D(Ab_1), D(Ab_2), \cdots, D(Ab_m))$. $D()$ means the function to compute the affinity.

Step3) Select n best individuals based on their affinities from the m original antibodies. These antibodies will be referred to as the elites.

Step4) Sort the n selected elites in n separate and distinct pools in ascending order. They will be referred to as the elite pools.

Step5) Clone the elites in the pool with a rate proportional to its fitness. The amount of clone generated for these antibodies is given by Eq.(1).

$$P_i = round(\frac{n-i}{n} \times Q) \tag{1}$$

i is the ordinal number of the elite pools, Q is a multiplying factor for determining the scope of the clone and $round()$ is the operator that rounds towards the closest integer. Then, we can obtain $\sum p_i$ antibodies as $((Ab_{1,1}, Ab_{1,2}, \cdots, Ab_{1,p_1}), \cdots, (Ab_{n,1}, Ab_{n,2}, \cdots, Ab_{n,p_n}))$.

Step6) Subject the clones in each pools through either hypermutation or receptor editing process. The mutation rates, P_{hm} for hypermutation and P_{re} for receptor editing given by Eq.(2) and Eq.(3), are inversely proportional to the fitness of the parent antibody,

$$P_{hm} = a/D() \tag{2}$$

$$P_{re} = (D() - a)/D() \tag{3}$$

, where $D()$ is the affinity of the current parent antibody and a is an appropriate numerous value.

Step 7) Determine the fittest individual B_i in each elite pool from amongst its mutated clones. The B_i is satisfied with the following equation.

$$D(B_i) = max(D(Ab_{i,1}, \cdots, Ab_{i,p_i})), i = 1, 2, \cdots, n \tag{4}$$

Step 8) Update the parent antibodies in each elite pool with the fittest individual of the clones and the probability $P(Ab_i, \rightarrow B_i)$ is according to the roles: if $D(Ab_i) < D(B_i)$ then $P = 1$, if $D(Ab_i) \geq D(B_i)$ then $P = 0$, otherwise $exp(\frac{D(B_i)-D(Ab_i)}{\alpha})$.

Step 9) Replace the worst c elite pools with new random antibodies once every k generations to introduce diversity and prevent the search from being trapped in local optima.

Step10) Determine if the maximum number of generation G_{max} to evolve is reached. If it is satisfied with this condition, it terminates and returns the best antibody. Otherwise, go to Step 4).

IV. CORONARY HEART DISEASE DATABASE[12]

A. An overview of Framingham Heart Study

The Framingham Heart Study was the first prospective study of cardiovascular disease. The study began in 1948 under the direction of the National Heart Institute (now known as the National Heart, Lung, and Blood Institute) in the United States. Participants were randomly sampled from the town of Framingham, Massachusetts. Examination of participants has taken place every two years and the cohort has been followed for morbidity and mortality over that time period.

Over the five decades, the Framingham Heart Study has provided valuable insights into the epidemiology and pathophysiology of CHD. The study identified constitutional and environmental factors associated with the development of CHD and established the concept of risk factors and their joint effects.

B. Six-year follow-up experience

The following factors are known as a major risk factor for CHD and used for estimating the individual risk of developing CHD within a ten-year time period[20]: old age, smoking, high blood pressure, high cholesterol, and diabetes mellitus. Most of the factors were mentioned first in the reports on six-year follow-up in the Framingham Heart Study[21],[22].

Fig.5 shows the six-year follow-up in the Framingham Heart Study. The cohort consisted of 5,127 men and women aged 30 to 59 who were initially free of CHD. At the study inception in 1948, they received a physical examination, laboratory tests, and a lifestyle interview. The cohort has been followed by means of biennial examinations including a detailed medical

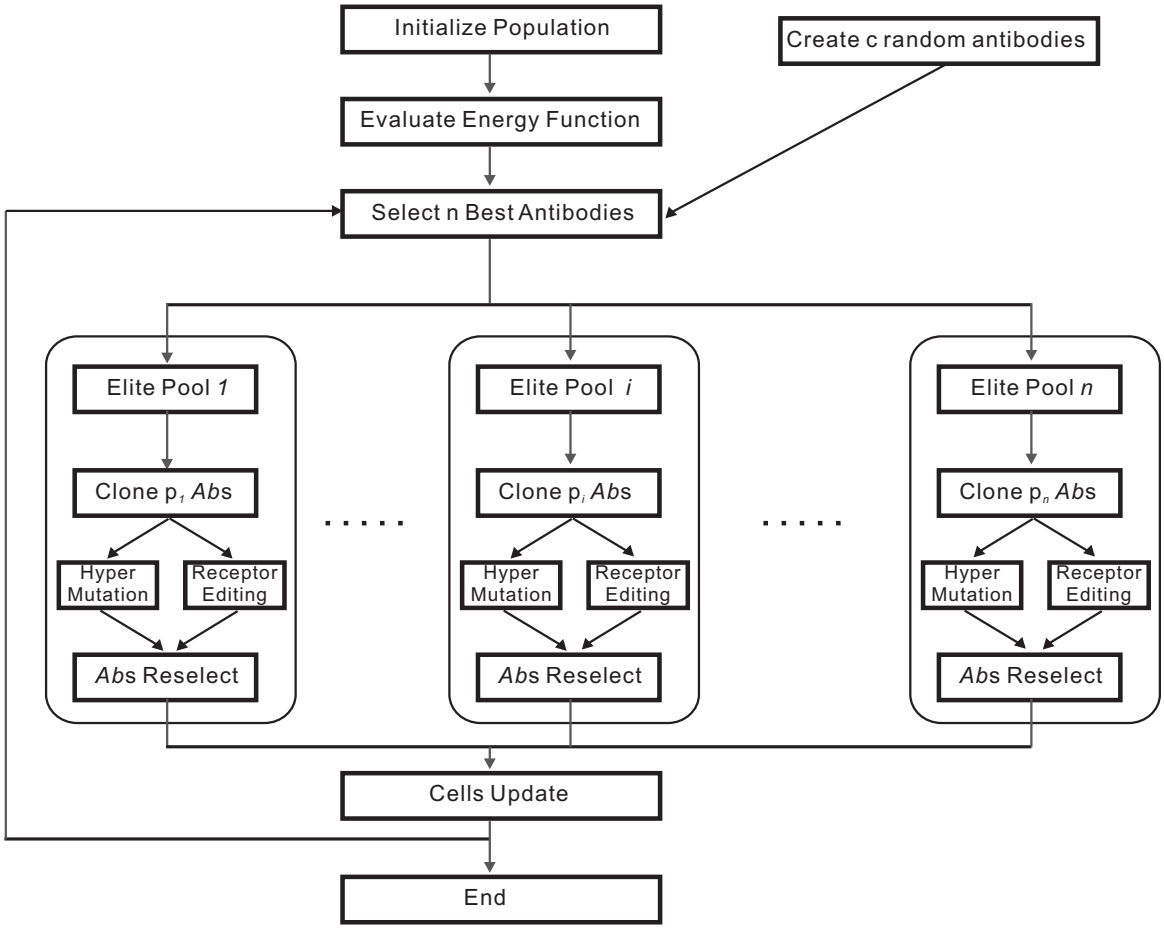

Fig. 4. A flowchart of RECSA model

history, a physical examination, and laboratory tests. There were 186 persons who developed CHD during the follow-up period. The six-year incidence of CHD in the age group of 45 or older was 9.1% in men and 4.5% in women.

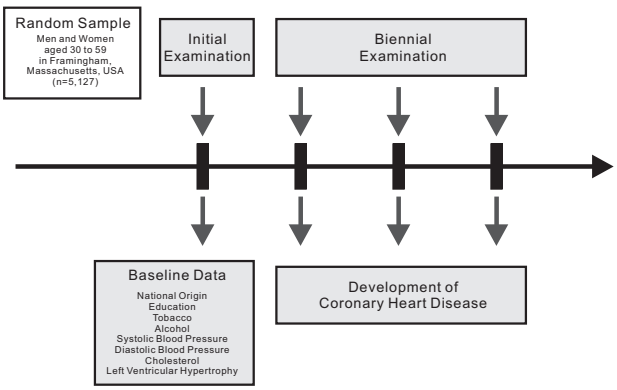

Fig. 5. Six-year follow-up in the Framingham Heart Study

C. Database Design

The CHD_DB is designed to reproduce the original data of the Framingham Heart Study. Requisite information is derived from the reports on six-year follow-up in the Framingham Heart Study [23],[24].

Table I shows the data items of the CHD_DB. Each of the datasets consists of ten data items: ID, development of CHD, and eight items that were collected from the initial examination (i.e. baseline data). The eight items; ORIGIN, EDUCATE, TOBACCO, ALCHOL, SBP, DBP, TC, and LVH, were examined whether it was associated with the development of CHD over six-year follow-up in the Framingham Heart Study. Using the CHD_DB, researchers will develop a prognostic system that will discriminate between those who developed CHD (CHD cases) and those who did not (Non-CHD cases) on the basis of the eight data items as shown in Fig.6.

The CHD_DB consists of four training datasets (Train_A, X, Y, and Z) and one testing dataset (Test) as shown in Fig.7. We previously reported that the developed prognostic system highly depended on the quality of training dataset [25]. The small proportion of cases to non-cases in the training dataset might contribute to a poor learning about the cases, and consequently, the developed prognostic system might have

TABLE I
DATA ITEMS OF CORONARY HEART DISEASE DATABASE

Data Item	Name	Value
ID	ID	Sequential Value
Development of CHD	CHD	0=No; 1= Yes
National Origin	ORIGIN	0=Native-born; 1=Foreign-born
Education	EDUCATE	0=Grade School; 1=High School, not graduate; 2=High School, graduate; 3=College
Tobacco	TOBACCO	0=Never; 1=Stopped; 2=Cigars or Pipes; 3=Cigarettes($<$20/day); 4=Cigarettes($\leq$20/day)
Alcohol	ALCOHOL	Continuous Value(oz/mo)
Systolic Blood Pressure	SBP	Continuous Value (mmHg)
Diastolic Blood Pressure	DBP	Continuous Value (mmHg)
Cholesterol	TC	Continuous Value (mg/dl)
Left Ventricular Hypertrophy	LVH	0=None; 1=Definite or Possible

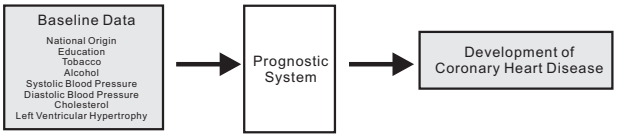

Fig. 6. Prognostic System for the development of CHD

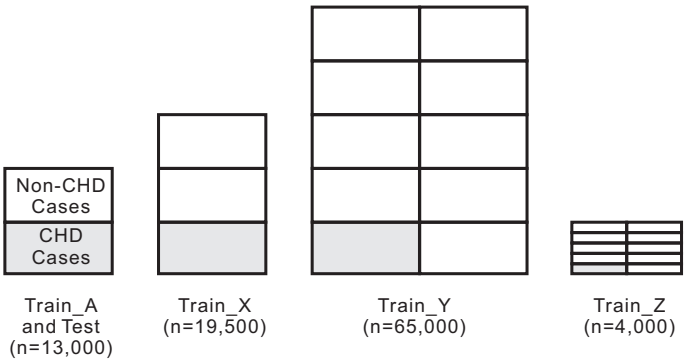

Fig. 7. Training and test datasets

lack of ability to identify potential cases from the population at risk. Therefore, the four training datasets are designed to have different proportions of CHD cases to Non-CHD cases. The number of combination of the eight data items is expected to be about six and a half thousand: 2 (ORIGIN) $\times$ 4 (EDUCATE) $\times$ 5 (TOBACCO) $\times$ 3 (ALCOHOL) $\times$ 3 (SBP) $\times$ 3 (DBP) $\times$ (TC) $\times$ 2 (LVH) = 6,480, where the numerical data are counted as three. In the original data of the Framingham Heart Study, the number of complete records was about four thousand and the proportion of CHD cases to Non-CHD cases in men aged 45 or older was about one to nine (i.e. the six-year incidence of CHD was 9.1%). Therefore, Train_A, Train_X, and Train_Y include 6,500 CHD cases and 6,500 ($\times$ 1), 13,000 ($\times$ 2), and 585,000 ($\times$ 9) Non-CHD cases, respectively, while Train_Z approximates to the original data both for the number of total records and the proportion of CHD cases to Non-CHD cases. On the other hand, for the testing dataset, we prepared 6,500 CHD cases and 6,500 Non-CHD cases separately from the training datasets.

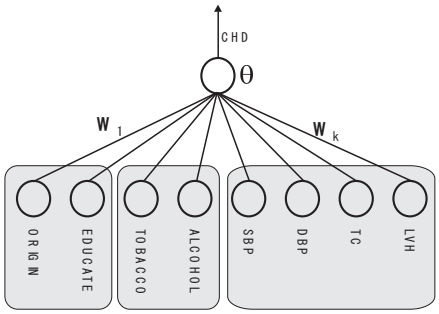

Fig. 8. The immune system for classification of CHD

V. EXPERIMENTAL RESULTS

First, we describes the preliminary requirements for the classification of CHD_DB by using RECSA model. The antibody consists of k paratopes that the part of the molecule of an antibody that binds to an antigen. There are $k = 8$ input signals in each record as shown in Table I, except ID and CHD. The antibody in the experimentation is assumed that the paratope consists of the weight, $\boldsymbol{W_k}$, for input signals and threshold value, θ, for an output, then the search for the optimal set of $P(w_1, w_2, \cdots, w_k, \theta)$ as shown in Fig.8 will be iterated by mutation. The initial set of $\boldsymbol{W}$ and θ is given an appropriate positive real number. In this work, $\boldsymbol{W}$ and θ were random numerous values in the range $[0, w]$ and $(0, \theta]$, where $w = 2.0, \theta = 8$.

In order to calculate the affinities, the agreement degree of antibodies and training data in CHD_DB is measured by using Eq.(5).

$$f(x) = \begin{cases} 1 & if \quad \sum_{i=1}^{k} w_i x_i \geq \theta \\ 0 & if \quad otherwise \end{cases} \tag{5}$$

If $f(x)$ is equal to a teach signal of training record, the antibody has a high affinity. The number of such antibodies shows an overall affinities in the classification of CHD_DB by using RECSA model.

Some parameters in RECSA model were taking same values described in [8]. That is, we use the parameters: $G_{max} = 500$, $m = 150$, $n = 100$, $Q = 50$, $\alpha = 100$, $k = 30$, $c = 10$, and $P_{hm} : P_{re} = 0.5 : 0.5$.

We use Java programming language on Pentium 4 3.20GHz with 1.50GB RAM. The version of Java is JDK1.6.0_07.The

TABLE II
NUMBER OF CORRECT ANTIBODIES(10 RUNS)

Dataset	Best	Worst	Mean	Variance	Time(sec)
Train_A	9,001(69.24%)	8,955(68.95%)	8976.8(69.05%)	361.8	588.17
Test	9,092(69.94%)	8,999(69.22%)	9042.3(69.56%)	866.4	-
Train_X	13,905(71.31%)	13,472(69.0%)	13680.4(70.16%)	21914.24	855.60
Test	9,081(69.85%)	9,008(69.29%)	9041.3(69.55%)	458.0	-
Train_Y	46,313(71.25%)	45,163(69.48%)	45948.0(70.69%)	198469.2	2691.23
Test	9,099(69.99%)	8,997(69.21%)	9058.8(69.68%)	1141.2	-
Train_Z	3,037(75.93%)	2,803(70.08%)	8923.6(68.64%)	2598.64	186.10
Test	9,064 (69.72%)	8,839(67.99%)	8944.2(68.80%)	4682.8	-

simulation for each dataset runs in 10 times by using Eclipse3.4. The experimentaion computed the best and worst antibody and mean of correct ratio in population.

Table II shows the classification results of Training dataset and Test dataset for 4 kinds of training datasets(Train_A, X, Y, and Z). The classification result by the adaptive learning method using neural networks in [27] was 82.3% for the test dataset of Train_A. The classification capability of our proposed method was not so good compared to that of [27]. However, there was no difference between the correct ratio for Training dataset and Test dataset over 10 runs, although each training dataset has different proportion rate of CHD cases to Non-CHD cases as shown in Fig.7.

VI. CONCLUSIVE DISCUSSION

The clonal selection principle established the idea that only those cells that recognize the antigens are selected to proliferate and differentiate. The method by [8], RECSA, explicitly takes into account the affinity maturation of the immune response by incorporating receptor editing method. This paper classified 4 kinds of Coronary Heart Disease Database. The computational results show about 69% correct ratio of test dataset. The RECSA model can classify the medical dataset, where includes some ambiguous data. However, the correct ratio is not higher than other classification methods[27], because the CHD database has some knowledge structure in the group of input terms as shown in Fig.8. Future works will focus the lateral interactive clonal selection algorithm [26] to enable the knowledge acquisition automatically.

REFERENCES

[1] L.N. de Castro and J. Timmis, “Artificial immune systems: A new computational Intelligence Approach,” Springer-Verlag(1996)

[2] L.N. de Castro and F.V. Zuben, “An evolutionary immune network for data clustering, ” Proc. IEEE SBRN, pp.84-89

[3] L.N. de Castro and F.V. Zuben, “Learning and optimization using clonal selection principle, ” IEEE Trans. Evolutionary Computation, Vol.6, No.3, pp.239-251(2002)

[4] J.E. Hunt and D.E. Ciijem “Learning using an artificial immune system, ” J. Network Comput. Applicat., Vol.19, No.2, pp.189-212(1996)

[5] D. Dasgupta, “Artificial immune systems and their applications, ” Springer-Verlag(1999)

[6] S.A. Hofmeyr and S. Forrest, “Immunity by design: An artificial immune system, ” Proc. of Genetic and Evolutionary Computation Conf., pp.1289-1296(1999)

[7] Y. Ishida, “The immune system as a prototype of autonomous decentralized systems: An overview,” Proc. Intl. Sympo. on Autonomous Decentralized Systems, pp.85-92(1997)

[8] S.Gao, H,Dai, G.Yang, and Z.Tang, “A novel clonal selection algorithm and its application to travelling salesman problem,” IEICE Trans. Fundamentals, Vol.E90-A, pp.2318-2325(2007)

[9] L.K. Verkoczy, A.S. Martensson, and D. Nemazee, “The scope of receptor editing and its association with autoimmunity,” Current Opinion in Immunology, Vol.16, pp.808-814(2004)

[10] R. Pelanda and R.M. Torres, “Receptor editing for better or for worse,” Current Opinion in Immunology, Vol.18, pp.164-190(2006)

[11] N.S. Longo and P.E. Lipsky, “Why do B cells mutate their immunoglobulin receptors? ,” TRENDS in Immunology, Vol.27, No.8, pp.374-380(2006)

[12] M. Suka, T. Ichimura and K. Yoshida, ”Development of coronary heart disease databases”, Proc. of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2004), Vol.2, pp.1081-1088(2004)

[13] F.M. Burnet, “Clonal selection and after,” in Theoretical Immunology, G.I. Bell, A.S. Perelson, and G.H. Pimbley Jr. Eds. New York; Marcel Dekker, pp.63-85(1978)

[14] F.M. Burnet, The Clonal Selection Theory of Acquired Immunity. Cambridge, U.K.: Cambridge Univ. Press(1959)

[15] M.C. Nussenzweig, “Immune receptor editing: Revise and select,” Cell, Vol.95, pp.875-878(1998)

[16] S. Tonegawa, “somatic generation of antibody diversity,” Nature, Vol.302, pp.575-581(1983)

[17] V. Kouskoff and D. Nemazee, “Role of receptor editing and revision in shaping the B and T lymphocyte repetorie,” Life Sciences, Vol.69, pp.1105-1113(2001)

[18] A.S. Perelson and G.F. Oster, “Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self-nonself discrimination,” J. Theor. Biol., Vol.81, No.4, pp. 645-670(1979)

[19] A.S. Perelson, “Immune network theory,” Immunological Review, Vol.110, pp.5-36(1993)

[20] P.J. Lucas and A. Abu-Hanna, “Prognostic methods in medicine,” Artif. Intell. Med. Vol.15, pp.105-119(1999)

[21] W. Penny and D. Frost, “Neural networks in clinical medicine,” Med. Decis. Making Vol.16, pp.386-398(1996)

[22] P.W.F. Wilson, R.B. D’Agostino, D. Levy, A.M. Belanger, et.al., “Prediction of coronary heart disease using risk factor categories,” Circulation Vol.97, pp.1837-1847(1998)

[23] T.R. Dawber, W.B. Kannel, N. Revotskie, J. 3rd. Stokes, et.al., “Some factors associated with the development of coronary heart disease: six year’s follow-up experience in the Framingham Study,” Am. J. Public Health, Vol.49, pp.1349-1356(1959)

[24] W.B. Kannel, T.R. Dawber, A. Kagan, N. Revotskie, et.al., “Factors of risk in the development of coronary heart disease: six-year follow-up experience,”, The Framingham Study. Ann. Intern. Med., Vol.55, pp.33-50(1961)

[25] M. Suka, S. Oeda, T. Ichimura, K. Yoshida, et.al., “Comparison of proportional hazard model and neural network models in a real data set of intensive care unit patients”, Proc.of MEDINFO2004(the World Congress on Medical Informatics), pp.741-745(2004)

[26] S.Gao, H,Dai, J.Zhang, and Z.Tang, “An expanded lateral interactive clonal selection algorithm and its application,” IEICE Trans. Fundamentals, Vol.E91-A, pp.2223-2231(2008)

[27] S. Oeda, T. Ichimura,and K.Yoshida, “Immune Multi Agent Neural Network and Its Application to Coronary Heart Disease Database”, Proc. of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2004), Vol.2, pp.1097-1105(2004)