# CVD level prediction processor using DNA computing

**Kyu-Yeul Wang**[a]**, Byung-Soo Kim, Chong-Ho Lee, Jaehyun Park, and Duck-Jin Chung**

*School of Information and Telecommunication Engineering, Inha University,*

*253 Younghyun-Dong, Nam-Gu, Incheon 402–751, Korea*

a) *wangky1007@inhaian.net*

**Abstract:** We propose a CVD (Cardiovascular disease) level prediction processor which has reduced computational efforts and time with insignificant diminution of CVD diagnosis accuracy when compared to the conventional CVD level prediction methods. The proposed DNA computing algorithm for implementation of CVD level prediction processor uses masks for simple nonspecific hybridization. Using the proposed mask, we can reduce HW computational efforts. Besides this, we decrease processing time resulting by co-processing hybridizations with a number of parallel matching units. The proposed CVD level prediction processor diagnoses CVD data of patients within 3.4 us. Therefore, the proposed CVD level prediction processor can be applied in a CVD level diagnosis system.

**Keywords:** cardiovascular disease, DNA computing, hypernetwork

**Classification:** Science and engineering for electronics

## References

[1] J.-H. Eom, S.-C. Kim, and B.-T. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2465–2479, May 2008.

[2] S. A. Tsaftaris, A. K. Katsaggelos, T. N. Pappas, and E. T. Papoutsakis, "How can DNA computing be applied to digital signal processing?," *IEEE Signal process. mag.*, vol. 21, no. 6, pp. 57–61, Nov. 2004.

[3] B.-T. Zhang, "Hypernetworks: A Molecular Evolutionary Architecture for Cognitive Learning and Memory," *IEEE Computational Interlligence Magazine*, vol. 3, no. 3, pp. 29–63, Aug. 2008.

[4] J. E. Bartlett, J. W. Kotrlik, and C. C. Higgins, "Organizational Research: Determining Appropriate Sample Size in Survey Research," *Information Technology, Learning, and Performance*, vol. 19, no. 1, pp. 43–50, 2001.

# 1 Introduction

Recent survey shows that the cardiovascular disease (CVD), which includes hearts disease and stroke, is one of the leading causes of the death without distinction of sex and age in the United States and all over the world [1]. But CVD diagnosis methods such as electrocardiography, sonography, blood test, angiography, and so on expend massive time and costs on experiments. To make final decision, a couple of tests should be carried out and it is unsafe and expensive. So, there are needs for novel CVD level prediction methods to compensate the defect and risk of previous medical CVD prediction methods. To solve the problems of medical CVD prediction methods, some studies proposed CVD level prediction method using machine learning or DNA computing [1]. These methods predict CVD level resulting from software based prediction schemes and spend lots of time to compute Maximum likelihood CVD level. In our study, we propose HW based CVD level prediction processor without reduction of prediction accuracy in reduced time. This article is organized as follows. In section II, we suggest proposed DNA computing algorithm for CVD level prediction processor. In section III, we describe the procedure of parameter decision for CVD level prediction processor design and then we explain the implementation of CVD level prediction processor in section IV. The implementation results and performance are elucidated in the section V. Finally, in the section VI, we make conclusion of this paper.

# 2 Proposed DNA computing algorithm for silicon based CVD level prediction processor

In the application like a CVD level prediction, the nonspecific hybridization is actually necessary, where an exact match is typically not possible due to the nature of the signals and noise [2] and it would guarantee accuracy as a disease diagnosis system. But it has massive parallel computation facts because nonspecific hybridization procedure needs a large number of executions for the stability of experiments. We need huge amount of memory to execute this maximum parallel operation. Moreover nonspecific hybridization operation needs identical number of random number generators. On the other hand, if this process is sequentially carried out, the prediction method spends lots of time for iterative operation. So, we need silicon based high performance processor for the CVD level prediction and we modified conventional DNA Computing algorithm for implementation of the proposed CVD level prediction processor.

## 2.1 Proposed DNA computing algorithm for memory reduction

The reference and test data are the amplified by the polymerase chain reaction (PCR) to be enforced at experiments and then amplified DNA molecules are hybridized in a hybridization temperature. The hybridization is the most computational process in the DNA computing for the CVD level prediction because this process operates hybridization between a test data of patient

and the reference data of previous diagnosis results. In the conventional DNA computing algorithm, following the amplification, the reference data and test data is stored in the memory. In the case of CVD prediction, we know that the reasonable prediction results are obtained with amplifying a data more than 1000 through empirical results. But this mechanism is not acceptable because the memory is limited in a general system. So, we used multiple hybridization modules and carried out hybridization iteratively to reduce the requirements of memory.

### 2.2 Proposed DNA computing algorithm for computational efforts

In the phase of computational efforts, the nonspecific hybridization needs random number generation. As a result, the computational efforts are increased exponentially to the amount of the CVD data. To reduce these computational efforts, we propose a mask for the hybridization. The mask represents the points that make not wobble but the watson-crick pairing. The masks are stored in the memory and are used for the hybridization, repetitively. Recursive use of the mask causes somewhat of decrease of accuracy because of an iterative wobble of the hybridized double strand. To prevent the reduction of accuracy, we need to get a set of orthogonal masks. So, we generated a mask and then discern whether the mask is existed in the generated set or not. If the generated mask is not existed in the mask set, the mask is added to the mask set. For the negligible accuracy reduction, we need to get a number of masks. This evolutionary learning process to find the maximum-likelihood parameters for the train data is expressed by following equation [3].

$$\ln P(D|W) = \sum_{n=1}^{N} \left\{ \left[ \sum_{k=1}^{K} \frac{1}{C(k)} \sum_{i_1,i_2\cdots,i_k} w_{i_1,i_2\cdots,i_k}^{(k)} x_{i_1}^{(n)} x_{i_2}^{(n)} \cdots x_{i_k}^{(n)} \right] - \ln Z(W) \right\}. \tag{1}$$

Using mask, the learning process is translated to a following equation.

$$\ln P(D|W) = \sum_{n=1}^{N} \left\{ \left[ \sum_{k=1}^{K} \frac{1}{M(k)} \sum_{j=1}^{M(k)} m_{i_1,i_2\cdots,i_k}^{(k)} x_{i_1}^{(n)} x_{i_2}^{(n)} \cdots x_{i_k}^{(n)} \right] - \ln Z(W) \right\}. \tag{2}$$

Where, the $M(k)$ represents the number of Masks of k-hypernetwork and $m_{i1,i2,\ldots ik}^{(k)}$ represents the structure and weights of hyperedge. The $M(k)$ is deduced by cochran's sample size formula for categorical data [4] and Bernoulli probability function.

$$Var[m_{i_1,i_2\cdots,i_k}^{(k)}] = k\left(1 - \frac{k}{K}\right). \tag{3}$$

$$M(k) = \frac{t^2 \cdot Var[m_{i_1,i_2\cdots,i_k}^{(k)}]}{d}. \tag{4}$$

Where, the $k$ and $K$ represent cardinality and data dimension, respectively. The $t$ is value for selected alpha level and Eq. (3) is variance of the K cardinality mask and the d of Eq. (4) is acceptable margin of error. We decided the $d$ as a 0.1 for the compatible simplicity and randomization. As a result, we can query each train data with 90% reliability.

## 3 Simulation for the parameter decision of CVD level prediction

In CVD level prediction using the DNA computing, the amplification amount of CVD data and nonspecific hybridization condition are very important. That is, the ratio of the watson-crick pairing in the CVD data hybridization and how many data are hybridized are critical condition. To extract this information, we simulated CVD level predictions with 135 CVD data. The CVD data extracted from aptamer bio-chip has 3000 features which has real-valued brightness. The data has massive data capacity and redundancy. For the feature selection, the dimension of CVD data set is reduced by applying analysis of variance (ANOVA) [1] and we selected the top 128 proteins according to their significance score to build final classifier inputs. We choose 20% of CVD data for the test data and the rest are used for train data. The simulation is accomplished 100 times and then the results are represented with arithmetic average. With this simulation, we can decide the cardinality ($k$) of the CVD data to 28.
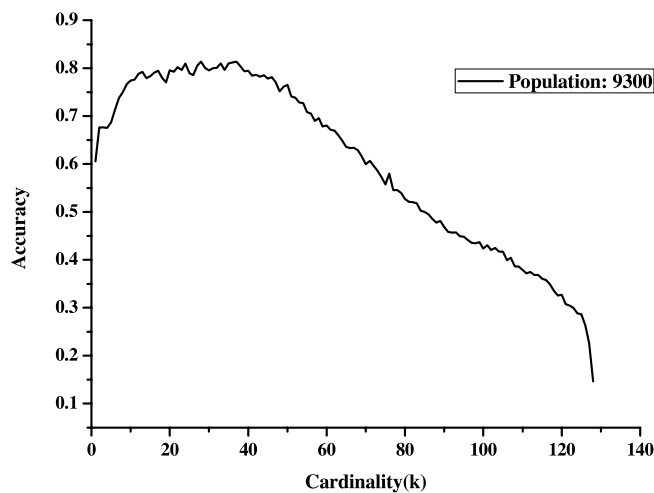


**Fig. 1.** Accuracy simulation results corresponding to Cardinality

## 4 Implementation of CVD level prediction processor

Our proposed processor consists of four processing path, controller, mask ROM, mask parser and prediction block. Each path is composed of CVD reference data block, matching units, accumulation block, and random mask ROMs. Each processing path processes CVD cases to discern the maximum likelihood prediction of CVD level, respectively. To do parallel processing, each path contains 30 matching units and a random mask ROM which offers masks to the matching unit for the nonspecific hybridization. The CVD processor's implementation and function will be explained at next paragraph.

## 4.1 Reference data block

The reference data is stored in the four RAMs according to the class of the CVD data to be compared with test data. We divided the CVD data into four classes. That is, the CVD data are constituted of normal data, stable angina data, unstable angina data, and myocardial angina data.

## 4.2 Matching Unit

The matching unit compares a reference data and a test data. The matching of CVD data is inexact matching. The matching unit generated output 1 if selected feature by masks are totally same. If not, the block sends output 0. The matching operation is carried out iteratively for the higher accuracy. The iterative operation can cover the various nonspecific hybridizations based on probabilistic occurrences. This mechanism represents nonspecific hybridization affected by hybridization temperature.

## 4.3 Controller and Accumulation block

The matching operation using nonspecific hybridization is adjusted by controller. The Controller assigns the reference data and test data to the matching unit according to the number of the required matching operation and generates an address of mask ROMs for the probabilistic nonspecific hybridization. Also, the controller enforced the enable signal to accumulate inexact matching results of matching units. The matching results are accumulated by CVD classes in the accumulation block. Finally, the these results are used for the CVD level prediction.
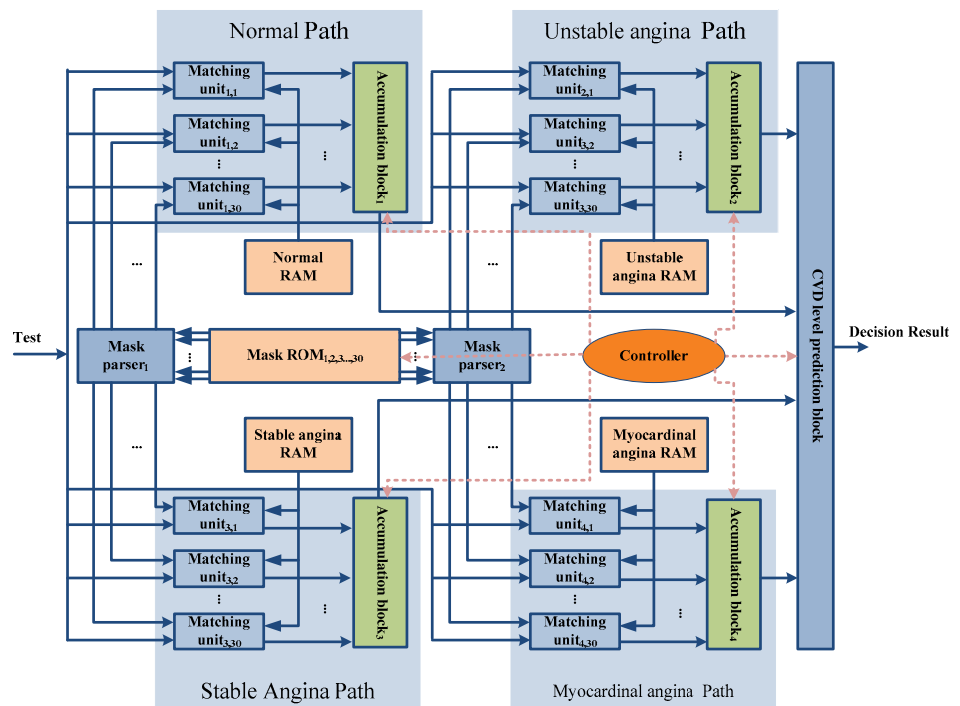


**Fig. 2.** Block diagram of CVD level prediction Processor

## 5 Implementation results

The proposed CVD level prediction processor is designed by using verilog HDL and it is implemented Magnachip 0.18 um CMOS process and FPGA. Considering that area of 2 input nand gate is normalized to 1, the proposed CVD level prediction processor has 154359 equivalent gate counts and 6.51 ns critical path without memory because the required memory depends on the reference data which is acquired previous medical data. The CVD level prediction processor operates at 153 MHz. Also, the processor is configured with Xilinx ISE using Virtex-5 xc5vlx330 and then the logic is tested with 100 MHz operating frequency. We can get the prediction results using test board within 3.4 us. The computational efforts are measured with regarding hybridization and amplification as a basic operation and then they are normalized with proposed CVD processor's computational efforts. The implementation results are compared with software based CVD prediction results in the Table I.

**Table I.** Comparison of SW based CVD level prediction and Proposed CVD level prediction processor

|  | SW based CVD prediction | Proposed CVD processor |
|---|---|---|
| Required Memory(KB) | 32140.8 | 150.5 |
| Processing Time for one test data | 357.14ms | 3.4us |
| Computational Efforts | 1.98 | 1 |
| Accuracy(%) | 81.3214 | 80.7143 |

The proposed CVD level prediction processor has compatible accuracy with a SW simulation and we can reduce the memory usage to about one hundredth by means of using modified algorithm. Using the masks eradicates the random number generation for nonspecific hybridization. So, the computational efforts are reduced by one halves. Moreover, the proposed processor reduced processing time to 3.4 us to classify one test data. The CVD level prediction program is executed by visual studio 2005. The simulation computer which executes the program has core 2quad CPU Q6600 at 2.4 GHz of Intel Corporation and 3.5 GB DDR2 main memory. The Windows XP is used as an operation system.

## 6 Conclusion

In this study, we proposed modified DNA computing algorithm for CVD level prediction processor. The modified algorithm required reduced memory requirements and computational efforts using by masks. The proposed processor is about 100,000 times faster than SW based CVD prediction and doesn't make accuracy reduction. Therefore, the proposed CVD level prediction processor can be applied in a CVD prediction method.