method and embedded evolutionary elements in Section 3. In Section 4, we explain data and experimental results. Finally, conclusions are suggested in Section 5.

II. Backgrounds

A. Hypernetwork Model

A hypernetwork is a generalized hypergraph which is represented with vertices and hyperedges. Fig.1 shows an example of hypernetwork. Formally, a hypernetwork is defined as $H = (V, E, W)$ where V, E, and W is a set of vertices, hyperedges, and weights, respectively [2].

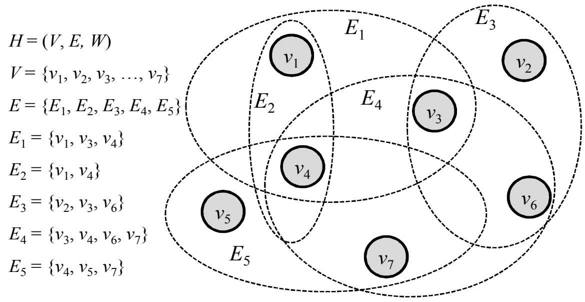


Fig. 1. An example of the hypernetwork. A hypernetwork consists of vertices and hyperedges. Hyperedges are connections among vertices. Unlike a conventional graph, hyperedges can contain more than three vertices.

A hyperedge of cardinality k, a hyperedge connected to k vertices, is referred to a hyperedge of order k or cardinality k. When orders of all hyperedges in a hypernetwork are k, we call it a k-uniform hypernetwork. Therefore, a conventional graph is considered as a 2-uniform hypernetwork with identical weights. Formally, a hypernetwork is defined to the probabilistic formula as followed. Given a data set $D = \{x^{(n)}\}_{n=1}^{N}$ of N samples, the hypernetwork represents the probability P

$$P(D \mid W) = \prod_{n=1}^{N} P(x^{(n)} \mid W) \quad , \tag{1}$$

$$P(x^{(n)} \mid W) = \frac{1}{Z(W)} \exp\left[\sum_{k=1}^{K} \frac{1}{C(k)} \times \sum_{i_1, i_1, \ldots, i_k} w_{i_1, i_1, \ldots, i_k}^{(k)} x_{i_1}^{(n)} x_{i_2}^{(n)} \ldots x_{i_k}^{(n)} \right] \tag{2}$$

, where $Z(W)$ denotes the normalizing term, and $C(k)$ is the number of possible k-order hyperedges [2]. Hypernetwork models are generated based on random sampling. Fig. 2 explains the creating procedure of a hypernetwork. In addition, Fig. 3 shows the process generating hyperedges. In Fig. 3, a hyperedges represented as simple rules are generated from a data sample and they become elements of a hypernetwork.

B. Brain and Cortical Thickness Data

Many previous MRI studies have used a manual region of interest or a fully automated voxel-based whole brain volumetric analysis to investigate and compare structural changes in the brain. However, those methods were insufficient for definitive representation of cortical structure. Recently, with the development of sophisticated 3-dimensional-based methods, measures of cortical thickness from MRI have been suggested to analyze cortical structure. Human cerebral cortical thickness has been used to quantify morphometric measures in the living brain. Therefore, the assessment of human cerebral cortical thickness has the importance in exploring normal brain development and in the determination of pathology, identifying regions where cortical thickness correlates with age, sex, mental disease state, genotype, and cognitive impairment.

Sex difference in the development of human brain is a topic of general interest. Many previous quantitative MRI studies of brain have reported sexual dimorphism in brain morphology. While gray matter (GM) and white matter (WM) size as well as total brain size are bigger in men than in women (Peters, 1991 [11], Goldstein et al., 2001 [12], Sowell et al., 2006 [13]), the relative proportions of GM volume to WM volume were higher in women (Allen et al., 2003 [14]) and higher GM concentration over widespread regions in women has also been reported (Good et al., 2001 [15], Luders et al., 2005 [16]). Recently, employing gender difference analysis of cortical thickness in healthy young adults with surface-based methods, Im *et al*. [10] have reported a significant localized cortical thickening in women, compared with men,

```
1. Given a data set D, divide the data into a training data set D_tr and a test
   set D_te.
2. Initialize a candidate hypernetwork H,
   H = (V,E,W) = (Ø, Ø, Ø)
3. Generate hyperedges to create H as follow
     E ← {}, V ← a set of attributes
     For i=1 to | D_tr |, where | D_tr | is the size of D_tr
       d_i ← the ith element of D_tr
       For j=1 to R, where R is sampling rate
         E_j ← Ø
         For m=1 to k, where k is a hyperedge's order
           v ← randomly selected attribute-value pair
           E_j ← E_j ∪ {v}
         End For
         E_j ← E_j ∪ {y_i}, where y_i is the class label of d_i
         W_j ← W', where W' is a initial weight value
         E ← E ∪ {E_j}, W ← {W_j}
       End For
     End For
     H ← (V, E, W)
```

Fig. 2. Pseudo-code for constructing a hypernetwork.

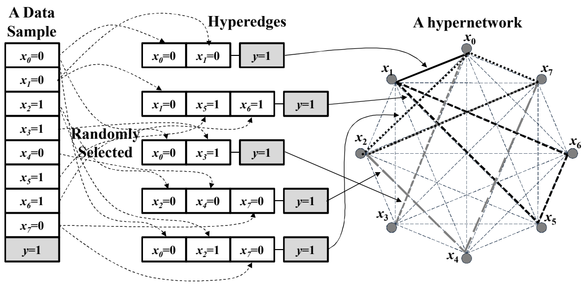


Fig. 3. Process for generating hyperedges. Hyperedges are built by random sampling of attributes.

in localized anatomical regions including frontal, parietal, and occipital lobes in most left hemisphere.

III. Evolving Hypernetworks and The Feature Selection Method

A hypernetwork represents a large combinatorial space since the number of combinations increases exponentially as the number of attributes increases. Learning in the hypernetwork is searching the large combinatorial spaces and evolutionary concepts are adapted to the model as a feasible searching strategy. That is, the model learning is conducted by the evolution of hypernetwork structure using diversity enforcement methods. Evolutionary learning methods enable the model to conserve and select significant attributes.

A. Replacement of hyperedges

No explicit cross-over operation is used in evolving a hypernetwork. Instead, the hypernetwork structure is varied by replacing hyperedges as mutation operation each learning epoch to enforce diversity of the model. Replaced hyperedges are determined by the weight of hyperedges. In this study, the victims are in either condition:

1) the weight is smaller than 0.1 × initial weight.
2) the hyperedge is not matched by any training sample.

Condition 1) is required to eliminate hyperedges which vote wrongly and condition 2) is needed because victims are useless in classifying the data in spite of the relatively large weight value. New hyperedges are regenerated to replace victims in next iteration.

B. Learning hypernetworks

After generating a hypernetwork, training and evaluating the model are conducted as learning where evolutionary methods are embedded. Training procedure of hypernetwork is based on equality comparison of hyperedges with training data [8]. That is, a vertex in hyperedges is compared with an attribute identical to the vertex in training data. When all vertices in a hyperedge are identical to attribute values in a training data sample, the sample is matched by the hyperedge. In matched cases, if the class label of a hyperedge is equal to one of a compared data sample, the weight of the hyperedge increases. Otherwise, the weight decreases. Finally, being similar to the training procedure, the evaluation is based on equality comparison. In addition, the classification is conducted with majority voting of hyperedges. Fig. 4 explains the training and classification procedure in the hypernetworks. After end of training, the model is evaluated with the test data set.

C. Extracting features from learned hypernetworks

With evolutionary learning, the fitter hyperedges survive only in a hypernetwork. Considering that a hyperedge is a combination of attributes, we can speculate that significant

$Y = \{y_1, \dots y_m\}$, where Y is a set of class labels
Training procedure of a hypernetwork
For i =1 to $|D|$, where $|D|$ is the size of a data set D
  $d_i \leftarrow$ the ith element of D
  For j=1 to $|E|$, where $|E|$ is the size of a hyperedges set $|E|$
    For k=1 to $|E_j|$, where $|E_j|$ is the order of E_j
      $v_k \leftarrow$ the kth vertex of E_j
      compare the value of v_k with one of v_k index in d_i
    End For
    If (matched)
      $y_e \leftarrow$ the class label of E_j, $y_d \leftarrow$ the class label of d_i
      If ($y_e = y_d$)
      Then $w_j \leftarrow w_j \times \alpha$
      Else $w_j \leftarrow w_j \times \beta$
    End If
  End For
End For

Classification procedure of a hypernetwork
$C = \{c_1, \dots, c_m\}$, where C is a set of counters of voted class labels
For i =1 to $|D|$, where $|D|$ is the size of a data set D
  $d_i \leftarrow$ the ith element of D
  For j=1 to $|E|$, where $|E|$ is the size of a hyperedges set $|E|$
    For k=1 to $|E_j|$, where $|E_j|$ is the order of E_j
      $v_k \leftarrow$ the kth vertex of E_j
      compare the value of v_k with one of v_k index in d_i
    End For
    If (matched)
      Increase c_{ye}, where y_e is the class label of E_j
    End If
  End For
  Classify y_n as the class label of d_i, where c = $\arg\max_n \frac{c_n}{\sum_{i=1}^{m} c_i}$
End For

Fig. 4. Training and classifying procedure for a hypernetwork.

attributes are likely to remain after learning. Therefore, learning in a hypernetwork enables to select features implicitly. In this study, we suggest an explicit method to select significant candidate attributes from a large number of attributes. The method is based on the occurrence frequencies of each attribute in survived hyperedges. To quantify the importance of attributes, we introduced an importance score for each attribute. The score reflects the contribution for each attribute to determine class labels of given data. Therefore attributes with higher score are selected as significant features. The importance score $s_{i,y}$ with respect to a class label y of the ith attribute is defined as:

$$s_{i,y} = \frac{F_x(freq(v_{i,x,y}))}{\sum_{x \in \{x|x \text{ is a value of } v\}} \sum_{y \in \{y|y \text{ is a class label}\}} freq(v_{i,x,y})} \quad (3)$$

In this formula, v denotes an attribute, x is the value of the attribute, *freq* is the occurrence frequency of the attribute in entire hyperedges, and F is an arbitrary function. The formula means the ratio of the occurrence frequency of a specific value and class label on each attribute in a hypernetwork. The bottom of formula represents a normalized term which means total occurrence frequencies. When the occurrence of an attribute is biased to a specific value for a class label, therefore, its score goes higher. On the contrary, when the values occur uniformly in hyperedges, the score becomes low.

$\arg\max_n \frac{c_n}{\sum_{i=1}^{m} c_i}$

IV. Experimental Results

A. Data preparation and experimental condition

In this study, we classified the gender based on cortical thickness measurements from MRI with the suggested hypernetwork (Fig. 5). The data represent cortical thickness of 52 healthy young adults including 31 men and 21 women. The data consist of measured values in the left and the right sphere of brain and each sphere has 40,962 thickness measured points. Hence, the data are represented with total 81,924 real-valued attributes. Detailed procedures to measure the cortical thickness are explained in Im et al.'s study [10]. Before learning with the hypernetwork, binarization of the data is required since the training of hypernetworks is based on equality comparison. Therefore, as preprocessing, the data were converted to binary type based on sample medians. For each attribute, that is, if the value of a sample is larger than the median of the attribute, the value becomes 1. Otherwise, the value turns 0. To evaluate the performance of the model, the data is divided into 70% training samples and 30% test samples. In training the data, we conducted a simple bootstrapping to balance the ratio of gender and overcome the shortage of training samples.

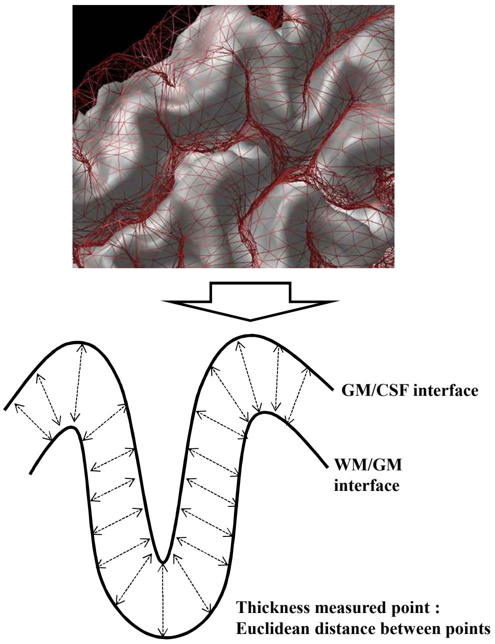


Fig. 5. Measuring the cortical thickness. Thickness values are obtained by evaluating Euclidean distances from GM/CSF boundary to WM/GM boundary surfaces with Constrained Laplacian based Automated Segmentation with Proximities (CLASP) algorithm [10].

Experimental conditions mean the determination of parameters of the hypernetwork model such as order of hyperedges, sampling rate, which is the number of generated hyperedges from a data sample, repeating numbers, weighting parameters. Table 1 shows the conditions in this study. In addition, we used Waikato Environmental for Knowledge Analysis (Weka) [17] to compare the classification accuracy with other machine learning methods.

TABLE I
EXPERIMENTAL CONDITIONS

Conditions	Hyperedge Order	Sampling Rate	Repeating Number	Weighting Parameter
Values	2~6 mixed	1000	20	Variable

The used hypernetwork models are mixed with 2 and 6 order hyperedges. That is, 1 ~ 3 attributes are selected from each sphere of the brain. Although weight-modifying parameter α is fixed to 1.001, β is variant according to learning progress. If Δ is less than 10%, β is 0.8. Otherwise, β is 0.5, where Δ is the difference of present training accuracy from previous one.

In this study, to select the significant measuring points, we defined the importance score $s_{i,y}$ to:

$$s_{i,y=m} = \frac{(freq(v_{i,x=0,y=f}) + freq(v_{i,x=1,y=m}))^2}{\sum_{x\in\{0,1\}} \sum_{y\in\{m,f\}} freq(v_{i,x,y})}, \tag{4}$$

$$s_{i,y=f} = \frac{(freq(v_{i,x=0,y=m}) + freq(v_{i,x=1,y=f}))^2}{\sum_{x\in\{0,1\}} \sum_{y\in\{m,f\}} freq(v_{i,x,y})}. \tag{5}$$

In above expressions, y denotes a class label, m means male, f means female, x_i denotes the value of v_i, which is the ith attribute, and $s_{i,y=m}$ denotes importance score related to male and $s_{i,y=f}$ represents one concerned with female. In these formula, we assume that the condition of 'x=0 and y=f' is equivalent to 'x=1 and y=m' because the value of x is 1 in male label and the value is 0 in female. Also, to enhance the effect of occurrence frequency, we adopted square function to F. To validate the proposed method, in addition, we compare the selected results and classification accuracy on the selected attributes with the selected results by gain-ratio feature selection method and accuracies by SVM feature selection, chi-square method as well as gain-ratio method.

B. Experimental results

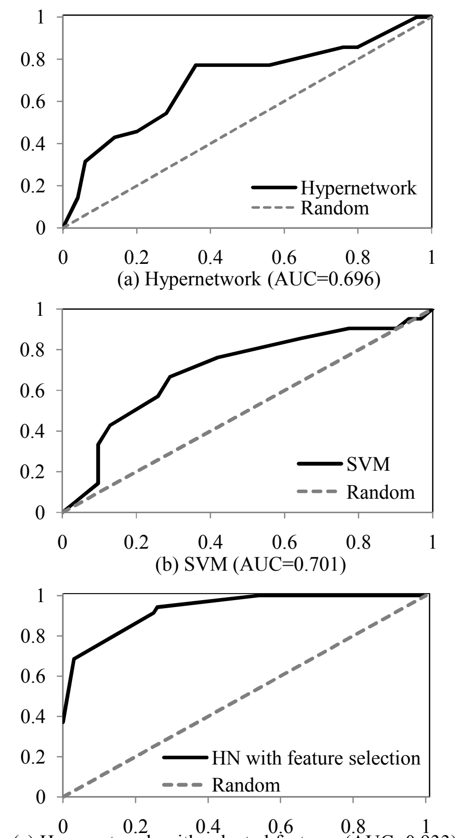
First, we compare the classification performance of the hypernetwork model with other machine learning algorithms without feature selection. Table 2 shows the comparison of evaluated results of the gender classification on the test data set. According to the results, all the accuracies are not very high because the data samples are too few compared with the

TABLE II
COMPARISON OF GENDER CLASSIFICATION PERFORMANCE

Methods	SVM	DT	k-NN	HN
Average Accuracy (%)	68.75	62.5	51.88	71.76
Standard Deviation	10.82	13.69	7.93	7.80

The values are accuracies on the test data set of machine learning methods. All values are averaged from the result of 10 learned hypernetworks. Linear kernel is used in SVM, k=1 in k-NN and DT means J48 decision tree in Weka. HN denotes a hypernetwork.

number of attributes. Especially, k-NN shows relatively lower accuracy since too many attributes may cause noise effects. Though, the hypernetwork has higher classification performances over other methods as 3~20%. We can guess the reason is that the hyperedges which consist of relatively significant attributes survive by learning. Although decision trees also use some parts of the attributes, they have the limitation in learning the patterns which are affected by complex relations among attributes. Dissimilar to decision trees, hypernetworks may be able to learn the complex patterns better since they are the ensembles of hyperedges. Fig. 6 shows ROC curves of the classification results based on female class of the hypernetwork and SVM. Although two models show a little different pattern in graphs, AUCs are similar to each other. Considering average accuracy, the variances of accuracy, and ROC curves, we can speculate that the hypernetwork has the competitive classification performance compared with other machine learning algorithms.



(a) Hypernetwork (AUC=0.696)

(b) SVM (AUC=0.701)

(c) Hypernetwork with selected features (AUC=0.933)

Fig. 6. ROC curves of classification results based on the female class of the hypernetwork, SVM and HN with feature selection. Each ROC curve is obtained from 10 different threshold values by averaging the results of 5 hypernetworks per threshold. The ROC curve and AUC of each method show that the feature selection method improves the classification performance.

Table 3 shows the candidates of significant measuring points of cortex thickness with the suggested feature selection method and its comparison with the list of attributes selected by gain-ratio method in Weka. We selected total 80 candidates of measuring points based on importance score. That is, 20 candidates with higher scores are selected in four cases of each brain sphere and gender. It is remarkable that most of high-ranked candidates are in male right sphere. To validate the effect of each feature selection, we classified gender with 80 selected features by the suggested method, gain-ratio, SVM feature selection, and chi-square method, separately. Table 4 represents the classification results of them. We can observe that the classification performance improves drastically with all feature selection methods according to Fig. 6 (c) and Table 4. Although SVM feature selection method shows the best accuracy performance, we can speculate that the proposed method shows as good performance as other methods according to Table 4.

TABLE III

SELECTED MEASURING POINTS OF CORTICAL THICKNESS

Male		Female	
Left	Right	Left	Right
36996 (106)	**40050 (2)**	**12001 (775)**	23414 (1116)
11332 (189)	**39943 (46)**	27332 (6336)	2157 (70945)
39949 (962)	**39936 (49)**	2166 (28105)	39246 (1047)
36962 (2146)	**40546 (20)**	12003 (1136)	10245 (79251)
38032 (64)	**40051 (1)**	33564 (19904)	33190 (60815)
38172 (475)	**40413 (117)**	38574 (16528)	34500 (59514)
9565 (957)	**40417 (54)**	12743 (38193)	9908 (1135)
36922 (1654)	**39948 (4)**	3015 (1004)	32880 (1102)
9270 (380)	**13121 (243)**	39584 (15213)	11447 (77141)
11760 (223)	**40506 (174)**	12007 (1074)	32183 (57182)
21459 (308)	**39937 (48)**	12753 (1059)	32824 (1119)
36913 (2175)	**9999 (25)**	20075 (34492)	26374 (43940)
38173 (951)	**39945 (36)**	38579 (1097)	2884 (71520)
11765 (307)	**40045 (28)**	12004 (1137)	17443 (72818)
29476 (5695)	**40409 (141)**	12016 (37296)	39240 (56635)
37975 (1278)	**2509 (169)**	34307 (20007)	32648 (57656)
40055 (3110)	40504 (919)	39627 (15096)	8257 (64795)
40368 (897)	**38172 (475)**	27337 (6312)	11499 (1030)
2959 (1959)	**40408 (143)**	3018 (1015)	9828 (995)
9533 (1185)	**16209 (41)**	12073 (37550)	32881 (1027)

The numbers indicate the indices of the measured points. The numbers in the parentheses are the ranking in the list of feature selection result based on gain-ratio. Bold numbers are top 1% of ranking of the gain-ratio result. The attributes are selected from 10 hypernetworks with 70% classification accuracy.

TABLE IV

CLASSIFICATION RESULTS WITH 80 SELECTED FEATURES

FS Method	Classifiers				
HN	SVM_L	SVM_Q	DT	k-NN	HN
Accuracy	**89.4**	**86.47**	**74.1**	**85.3**	**87.6**
Stdev	6.08	7.87	12.8	6.36	4.34
GR	SVM_L	SVM_Q	DT	k-NN	HN
Accuracy	81.2	75.88	77.1	81.2	85.9
Stdev	10.5	10.16	10.0	6.34	5.39
SVM	SVM_L	SVM_Q	DT	k-NN	HN
Accuracy	100	100	75.94	100	100
Stdev	0	0	10.0	0	0
X^2	SVM_L	SVM_Q	DT	k-NN	HN
Accuracy	90.59	91.77	76.47	85.88	88.82
Stdev	11.09	4.11	3.04	4.96	3.34

The table shows the gender classification results with 80 selected attributes by the hypernetwork-based method and gain-ratio method, respectively. In the table, HN denotes the hypernetwork-based feature selection method and GR means gain-ratio method. SVM_L is SVM with linear kernel and SVM_Q means one with quadratic kernel. The

TABLE V
CONFUSION MATRIX OF HYPERNETWORKS WITH AND WITHOUT FEATURE SELECTION

		Actual gender			
		All attributes		Selected attributes	
	Predicted	Male	Female	Male	Female
Precision (%)	Male	**72.03**	27.97	**90.72**	9.28
	Female	28.85	**71.15**	16.44	**83.56**
Recall (%)	Male	**85.0**	47.15	**88.0**	12.86
	Female	15.0	**52.85**	12.0	**87.14**

The columns denote actual genders and the rows are predicted genders. The values are averages from 10 times evaluations.

Finally, Table 5 presents the confusion matrix with and without feature selection by the suggested method. We can observe that the feature selection method improved the precision and the recall remarkably. Especially, the recall of female is improved by about 35%. It means that more hyperedges with male class label survive than female hyperedges without feature selection. Considering that the size of two class labels are balanced in training set, we can speculate that too many features have noise effects on classifying the gender and the effects are stronger in female cases. We can also guess that the noise effects are weakened by feature selection.

V. CONCLUSIONS

In this study, we proposed a novel feature selection method based on evolutionary hypernetworks and applied the method to classify the gender based on cortical thickness of healthy young adults from MRI. The experimental results present the suggested feature selection method can improve the classification performance drastically. In addition, the results present the hypernetwork model shows good performance compared with other machine learning methods with and without feature selection.

Since the hypernetwork model is to try to search much larger solution spaces with large number of comparison calculations, however, the model has heavier time costs than other machine learning methods. Therefore, efficient comparison methods and parallel processing are required to overcome the limitation.

With respect to brain researches, the selected measured points can be mapped to brain images. Therefore, by analyzing the mapped images, it is important to discover relations between the selected measure points and the function of brain.

ACKNOWLEDGMENT

Jung-Woo Ha thanks Kwon-Il Kim, Jun-Sik Kim, Sun Kim, Jae-Hong Eom, Min-Su Lee and Myung-ran Ha for discussion

REFERENCES

[1] D. Zhou, J. Huang, and B. Schoelkopf, "Learning with hypergraphs: Clustering, classification, and embedding", *Advances in Neural Information Processing Systems* (*NIPS*) 19. 2007.

[2] B. -T. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory", *IEEE Computational Intelligence Magazine*, 3(3), pp. 49-63, 2008.

[3] B. -T. Zhang and H. -Y. Jang, "Molecular programming: evolving genetic programs in a test tube", *The Genetic and Evolutionary Computation Conference* (*GECCO* 2005), vol. 2, pp. 1761-1768, 2005.

[4] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, 266(5187), pp. 1021–1024, 1994.

[5] B. -T. Zhang and J. -K. Kim, "DNA hypernetworks for information storage and retrieval", *Lecture Notes in Computer Science*, *DNA* 12, 4287 (2006), pp. 298-307, 2006.

[6] S. Kim, M. -O. Heo, and B. -T. Zhang, "Text classifiers evolved on a simulated DNA computer", *IEEE Congress on Evolutionary Computation* (*CEC* 2006), pp. 9196-9202, 2006.

[7] J. -K. Kim and B. -T. Zhang, "Evolving hypernetworks for pattern classification", *IEEE Congress on Evolutionary Computation* (*CEC 2007*), pp.1856~1862, 2007.

[8] J. -W. Ha, J. -H. Eom, S. -C. Kim and B. -T. Zhang, "Evolutionary hypernetwork models for aptamer-based cardiovascular disease diagnosis", *Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation* (2007), pp. 2709-2716, 2007.

[9] C. -H. Park, S. -J. Kim, S. Kim, D. -Y. Cho and B. -T. Zhang, "Finding cancer-related gene combinations using a molecular evolutionary algorithm", *IEEE 7th international conference on Bioinformatics & BioEngineering* (*BIBE 2007*), pp. 158-163, 2007.

[10] K. H. Im, J. -M. Lee, J. K. Lee, Y. -W. Shin, I. Y. Kim, J. S. Kwon and S. I. Kim, "Gender difference analysis of cortical thickness in healthy young adults with surface-based methods", *Neuroimage* 31(2006), pp. 31-38, 2006.

[11] M. Peters, "Sex differences in human brain size and the general meaning of differences in brain size". *Can. J. Psychol.* 45, pp. 507– 522, 1991.

[12] J. M. Goldstein, L. J. Seidman, N. J. Horton, N. Makris, D. N. Kennedy, V. S. Caviness Jr., S. V. Faraone, and M. T. Tsuang, "Normal sexual dimorphism of the adult human brain assessed by in vivo magnetic resonance imaging", *Cereb. Cortex* 11, 490– 497, 2001.

[13] E. R. Sowell, B. S. Peterson, E. Kan, R. P. Woods, and J. Yoshii, "Sex Differences in Cortical Thickness Mapped in 176 Healthy Individuals between 7 and 87 Years of Age", *Cerebral Cortex* 2007 17(7), pp. 1550 – 1560, 2006.

[14] J. S. Allen, H. Damasio, T. J. Grabowski, J. Bruss, and W. Zhang, "Sexual dimorphism and asymmetries in the gray– white composition of the human cerebrum", *NeuroImage* 18, pp. 880–894, 2003.

[15] C. D. Good, I. Johnsrude, J. Ashburner, R. N. Henson, and K. J. Friston, "Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains", *Neuroimage* 14, pp. 685–700, 2001.

[16] E. Luders, K. L. Narr, P. M. Thompson, R. P. Woods, and D. E. Rex DE, "Mapping cortical gray matter in the young adult brain: effects of gender", *Neuroimage* 26, pp. 493–501, 2005.

[17] University of Waikato New Zealand, "Waikato Environment for Knowledge Analysis (Weka) Version 3.4.11", Available: http://www.cs.waikato.ac.nz/ml/weka/index.html