Chapter 7

Experimental Results

"No amount of experimentation can ever prove me right, a single experiment can prove me wrong"

- Albert Einstein

This chapter presents experimental results to analyse the behavior of the GAIS model on different classification problems. The performance of GAIS is investigated under different control parameter values. The data sets were collected from the UCI Machine Learning Repository [7]. The patterns in each data set were discretised and converted to binary strings (as explained in section 6.2.1). For each experiment, one of the classes in a data set is selected as the self set. The self set is then used to train the ALCs with the adapted negative selection and the positive selection methods. The other classes in the data set represent the non-self patterns.

All experiments used a 30-fold cross validation self set. The self set was randomly divided into thirty disjoint sets. The ALCs were trained on 29 of these self sets and were tested with a test set that consisted of the remaining self set (the training set that was left out during training) and the unseen non-self patterns. In each experiment the initial population size in the GA was set to 100 chromosomes ($I = 100$) and the rate of elitism was set to 30% ($e = 0.3$). The window size in the GA and GAIS was set to 4.0 ($WindowSize = 4.0$) and $\mu_T = 0.01$ to test for convergence in the population and ALC set respectively. These values were found empirically to deliver good performance. All experimental results in this chapter are averages over 30 simulations with the selected self class in *italic* print. The best parameter settings are printed in **bold** in each table.

The results in each table, starting with the leftmost column, are the iteration size (IS) and param-

eter $w_1 * 100\%$ (W1) in the fitness function of the GA (recall that $w_2 = 1.0 - w_1$). The following results are the averages after the ALC set in GAIS has converged and all patterns were classified: the average number of ALCs in the active set (#ALCs), the average number of ALCs in the active set with memory status (#MemALCs), the average number of false positives (#fPos), the average number of false negatives (#fNeg), the average affinity distance threshold in the active set (ADT) and the average hamming distance between ALCs in the active set (HD). The average HD indicates the average hamming seperation among ALCs in the active set to cover non-self space. A higher value of HD indicates less overlap. The standard deviation is given in parentheses. A pattern from the self class that is falsely classified as non-self is referred to as a *false positive*. A pattern that is not from the self class and is falsely classified as self is referred to as a *false negative*. The average number of misclassified patterns for each parameter setting in each table is calculated as follows:

$$\#Misclassified = falsePositives + falseNegatives$$

Note that the accuracy is the average over all iterations. The interval-values for IS were calculated by

$$IS = \frac{is}{100} * \text{Size of data set}, is \in [25, 50, 75, 100]$$

The total number of iterations that GAIS executes is therefore equivalent to $\frac{\text{Size of the data set}}{IS}$. The selected values for w_1 were calculated by

$$w_1 = \frac{W1}{100}, W1 \in [25, 50, 75, 100]$$

Some of the tables are accompanied by figures that illustrate the results with the best parameter settings. These figures are the average number of ALCs in the active set, the average fitness of the ALCs in the active set and/or the average number of misclassified patterns at each iteration of GAIS. The figure that illustrates the average number of misclassified patterns, illustrates the average number of misclassified patterns after classifying IS patterns at each iteration. The number of patterns that was classified in the last iteration is less or equal to IS, i.e. the remaining patterns in the test set. The sum of the average number of misclassified patterns per iteration will correspond to the calculated number of misclassifications for the specific parameter settings in the corresponding tables.

Section 7.1 to section 7.5 discuss the results obtained from the GAIS model to classify the classes

in the Iris data set, Wisconsin breast cancer data set, Mushroom data set, Glass data set, and the Car evaluation data set respectively with different parameter settings. The results of each of these data sets are concluded in their respective sections and a comparison with C4.5 is given in section 7.6.

7.1 Iris

The iris data set contains three classes of fifty instances each, where each class refers to a type of iris plant. The setosa class is linearly separable from the versicolor class and the virginica class. The versicolor class and the virginica class are not linearly separable. The dataset consists of 150 patterns, evenly distributed among the three classes (33.3% each). Each pattern consists of four continuously valued attributes. The patterns were converted to binary strings of length 20.

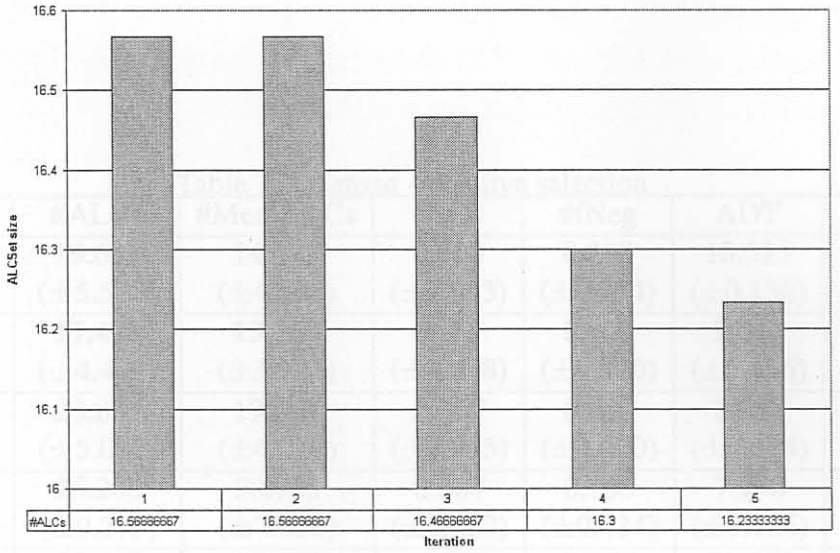
7.1.1 Setosa

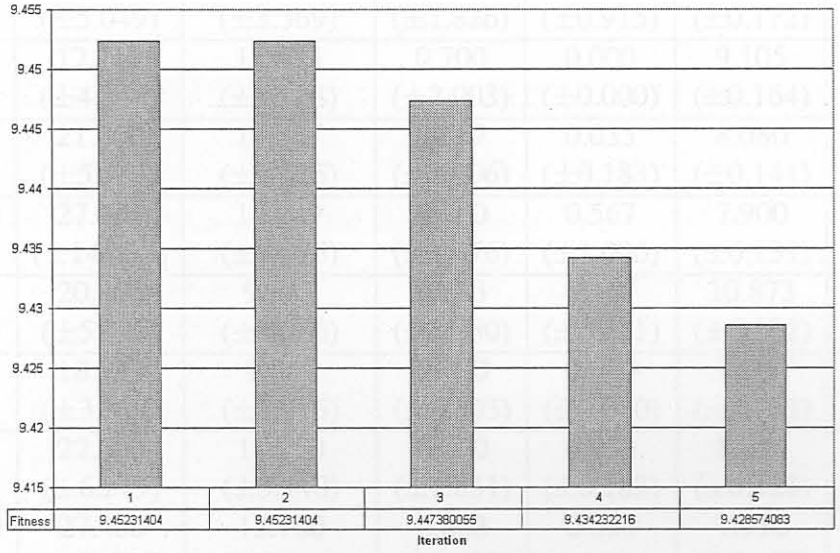
As a first experiment, setosa was selected as self. The results for training the ALCs with negative selection on patterns of the setosa class as self after convergence are summarised in table 7.1. Most of the parameter settings in table 7.1 had an average false negative classification of 0.000. The overall best result among the parameter settings with 0.000 false negative classification, is with IS=37 and W1=50 since the average number of ALCs in the active set of ALCs (#ALCs = 16.427) and the false positive classification (fPos = 0.500) are the lowest. This gives a number of 0.500 patterns misclassified ($\#Misclassified = 0.500 + 0.000 = 0.500$) and a correct classification of 99.666%. The number of ALCs with memory status in the active set was on average 12.980. Figure 7.1 shows that the active set of ALCs started with an initial average size of 16.566 ALCs in the active set and then decreased over five iterations to an average number of 16.233 ALCs in the active set. The average fitness of the ALC set decreased over the iterations from 9.452 to 9.428. There was a decrease in misclassification over the iterations with an increase at iteration four.

Table 7.2 summarises the results for training the ALCs with positive selection on patterns from the setosa class as self. The positive selection also has different parameter settings for which a false negative classification of 0.000 was obtained. When IS=37 and W1=50 the lowest misclassification was achieved ($\#Misclassified = 0.567 + 0.000 = 0.567$) with the least average number of ALCs in the active set of ALCs (#ALCs = 17.420). This gives a correct classification of 99.622%. The average number of ALCs with memory status in the active set of ALCs

Table 7.1: *Setosa* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	20.067 (±6.285)	14.727 (±5.004)	0.767 (±1.995)	0.167 (±0.531)	9.423 (±0.139)	9.393 (±0.213)
	50	**16.427 (±5.187)**	**12.980 (±4.203)**	**0.500 (±1.306)**	**0.000 (±0.000)**	**10.874 (±0.180)**	**8.012 (±0.188)**
	75	21.900 (±4.664)	17.520 (±3.731)	0.633 (±2.189)	0.033 (±0.183)	11.946 (±0.131)	6.138 (±0.253)
	100	26.667 (±9.932)	21.333 (±7.946)	0.567 (±1.478)	0.133 (±0.434)	12.099 (±0.110)	5.191 (±0.311)
74	25	19.511 (±5.504)	12.000 (±3.614)	0.767 (±1.995)	0.233 (±0.626)	9.328 (±0.167)	9.518 (±0.097)
	50	17.844 (±3.963)	11.711 (±2.652)	0.633 (±1.829)	0.000 (±0.000)	10.929 (±0.200)	8.011 (±0.191)
	75	23.567 (±5.117)	15.711 (±3.411)	0.667 (±1.826)	0.000 (±0.000)	11.946 (±0.178)	6.164 (±0.273)
	100	23.733 (±12.273)	15.822 (±8.182)	0.633 (±2.008)	0.567 (±1.135)	12.102 (±0.123)	5.148 (±0.377)
112	25	20.933 (±6.147)	9.883 (±3.183)	0.733 (±1.999)	0.300 (±0.702)	9.182 (±0.215)	9.616 (±0.057)
	50	16.933 (±4.017)	8.400 (±2.061)	0.567 (±1.478)	0.000 (±0.000)	10.878 (±0.163)	8.035 (±0.178)
	75	23.233 (±6.207)	11.617 (±3.104)	0.767 (±2.176)	0.000 (±0.000)	11.946 (±0.139)	6.131 (±0.209)
	100	27.400 (±10.627)	13.700 (±5.314)	0.500 (±1.480)	0.267 (±1.048)	12.107 (±0.172)	5.187 (±0.354)
150	25	20.567 (±4.636)	0.000 (±0.000)	0.733 (±1.999)	0.067 (±0.254)	9.193 (±0.131)	9.618 (±0.052)
	50	17.433 (±3.540)	0.000 (±0.000)	0.733 (±2.363)	0.000 (±0.000)	10.895 (±0.184)	8.013 (±0.173)
	75	21.667 (±5.616)	0.000 (±0.000)	0.700 (±1.643)	0.000 (±0.000)	11.956 (±0.154)	6.138 (±0.268)
	100	23.433 (±11.820)	0.000 (±0.000)	0.467 (±1.479)	0.400 (±0.724)	12.083 (±0.129)	5.151 (±0.352)





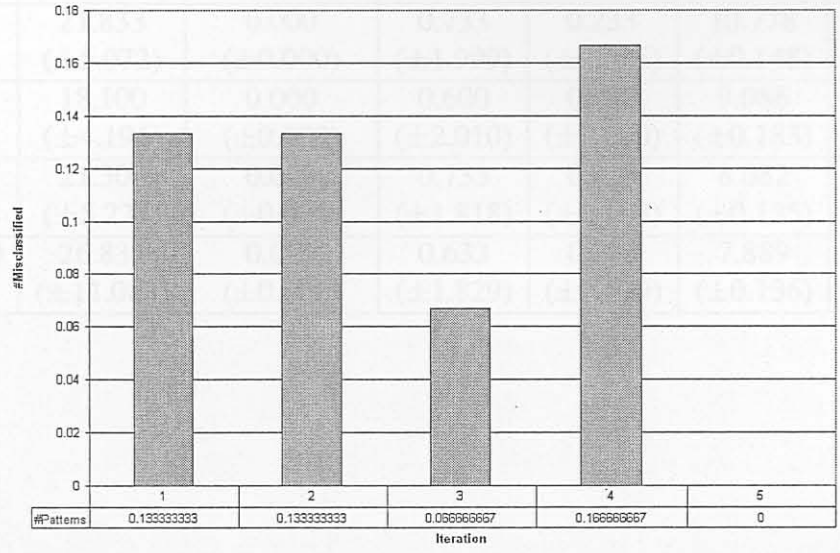


Figure 7.1: *Setosa* - Negative selection with IS=37 and W1=50

Table 7.2: *Setosa* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	19.653 (±5.572)	14.433 (±4.506)	0.700 (±1.643)	0.233 (±0.774)	10.525 (±0.136)	9.387 (±0.176)
	50	**17.420 (±4.409)**	**13.760 (±3.524)**	**0.567 (±1.478)**	**0.000 (±0.000)**	**9.043 (±0.196)**	**7.948 (±0.209)**
	75	23.833 (±5.093)	19.067 (±4.074)	0.767 (±1.995)	0.000 (±0.000)	8.087 (±0.154)	6.212 (±0.277)
	100	26.200 (±9.279)	20.953 (±7.426)	0.567 (±1.832)	0.133 (±0.434)	7.900 (±0.126)	5.183 (±0.384)
74	25	21.667 (±5.049)	13.578 (±3.369)	0.667 (±1.826)	0.300 (±0.915)	10.662 (±0.172)	9.546 (±0.069)
	50	17.711 (±4.846)	11.633 (±3.308)	0.700 (±2.003)	0.000 (±0.000)	9.105 (±0.164)	8.038 (±0.121)
	75	21.700 (±5.453)	14.467 (±3.635)	0.667 (±2.006)	0.033 (±0.183)	8.080 (±0.141)	6.214 (±0.274)
	100	27.133 (±14.063)	18.089 (±9.375)	0.400 (±1.476)	0.567 (±1.006)	7.900 (±0.131)	5.044 (±0.468)
112	25	20.400 (±5.090)	9.533 (±2.566)	0.833 (±2.350)	0.167 (±0.531)	10.873 (±0.192)	9.641 (±0.047)
	50	18.333 (±3.889)	9.067 (±1.915)	0.700 (±2.003)	0.000 (±0.000)	9.102 (±0.153)	8.037 (±0.185)
	75	22.500 (±6.279)	11.250 (±3.140)	0.600 (±1.831)	0.033 (±0.183)	8.071 (±0.128)	6.085 (±0.303)
	100	27.400 (±11.828)	13.700 (±5.914)	0.633 (±1.650)	0.333 (±1.124)	7.910 (±0.137)	5.236 (±0.311)
150	25	21.833 (±5.072)	0.000 (±0.000)	0.733 (±1.999)	0.233 (±0.626)	10.778 (±0.148)	9.615 (±0.042)
	50	18.100 (±4.196)	0.000 (±0.000)	0.600 (±2.010)	0.000 (±0.000)	9.086 (±0.183)	8.060 (±0.203)
	75	21.300 (±5.227)	0.000 (±0.000)	0.733 (±1.818)	0.033 (±0.183)	8.082 (±0.125)	6.187 (±0.218)
	100	26.833 (±11.083)	0.000 (±0.000)	0.633 (±1.829)	0.233 (±0.679)	7.889 (±0.136)	5.136 (±0.313)

is 13.760. Figure 7.2 shows that the active set of ALCs had an initial average size of 17.6 and decreased over five iterations to an average number of 17.266. The average fitness also decreased over the iterations from 9.458 to 9.446 and the average misclassification increased from iteration one to two, and then decreased to iteration five.

7.1.2 Versicolor

The classification results for training the ALCs with negative selection on patterns of the versicolor class as self are shown in table 7.3. The lowest misclassification ($\#Misclassified = 0.967 + 3.333 = 4.300$) was achieved with IS=37 and W1=75. This gives a correct classification of 97.133% with an average number of 21.567 ALCs in the active set of ALCs. The average number of ALCs with memory status in the active set was 17.220. Figure 7.3 shows that their was no change in the size of the active set over all iterations and the constant size is an average number of 21.566 ALCs. The average fitness of the ALC set was constant at 9.908 and the average number of misclassification increased from iteration one to iteration three and then decreased to iteration five.

The best classification result shown in table 7.4, is achieved with IS=37 and W1=50 when training the ALCs with positive selection on the patterns of the versicolor as self. The misclassification of 4.5 patterns ($\#Misclassified = 1.167 + 3.333 = 4.5$) gives a correct classification of 97.000%. The average number of ALCs in the active set of ALCs was 20.333 and an average number of 16.253 of these had memory status. Figure 7.4 shows that the average number of ALCs in the active set of ALCs was constant at 20.333 over all iterations. The average number of misclassification increased from iteration one to three and then decreased to iteration five.

7.1.3 Virginica

Table 7.5 shows that with IS=150 and W1=75 the lowest misclassification of 8.100 patterns ($\#Misclassified = 1.200 + 6.900 = 8.100$) was achieved when training the ALCs with negative selection on patterns of the virginica class as self. This gives a correct classification of 94.600% with an average number of 23.533 ALCs in the active set. An average number of 0.000 of the ALCs in the active set had memory status, since $IS = 150$ (which is the size of the data set) which implies that there was only one iteration. Since an ALC's status is evaluated after an iteration, the active set of ALCs in the first and only iteration cannot contain ALCs with memory status.

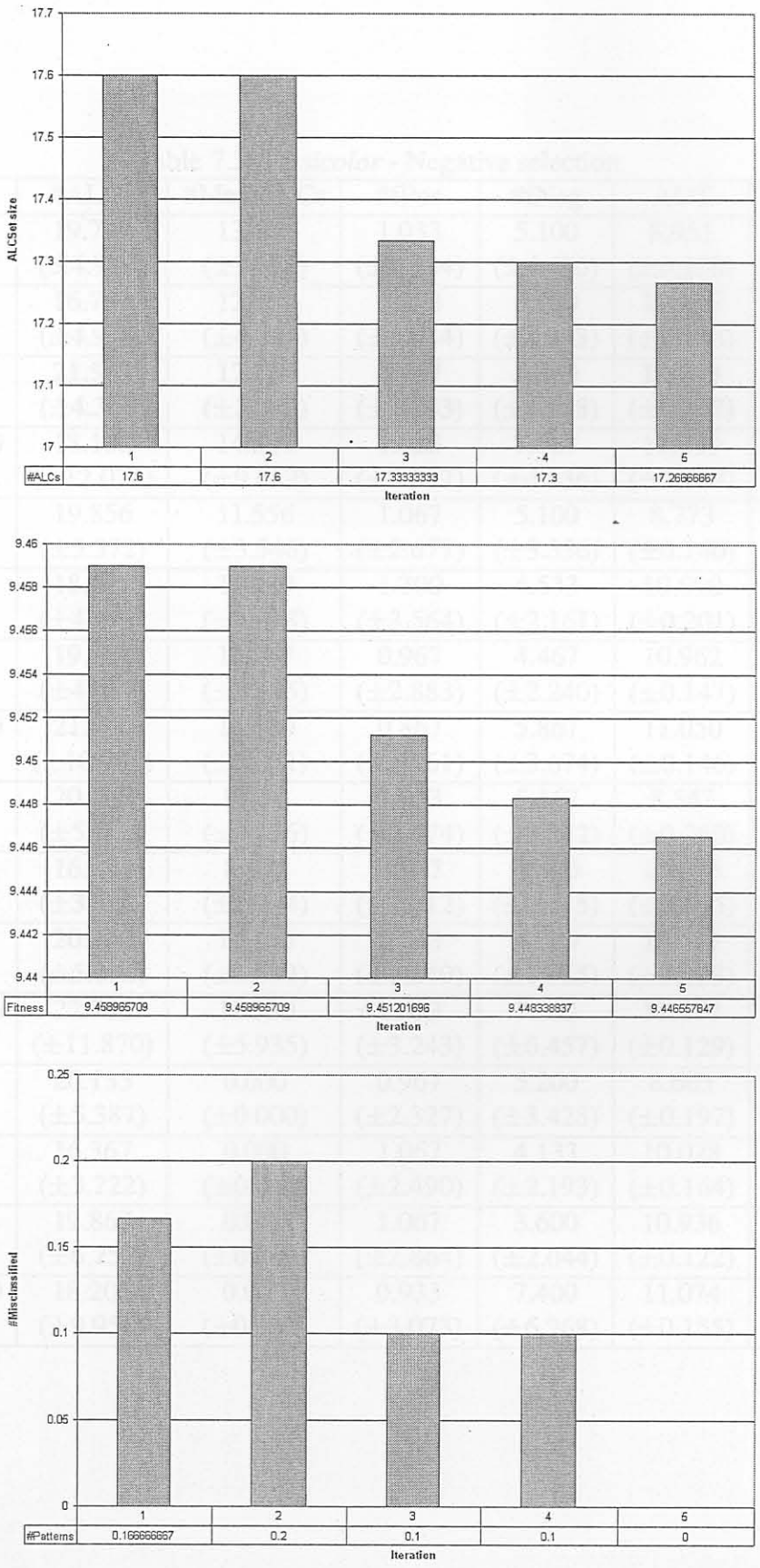


Figure 7.2: *Setosa* - Positive selection with IS=37 and W1=50

Table 7.3: *Versicolor* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	19.747 (±4.979)	13.660 (±3.953)	1.033 (±2.684)	5.100 (±4.436)	8.951 (±0.158)	9.454 (±0.148)
	50	16.753 (±4.979)	12.887 (±4.001)	1.233 (±3.014)	5.000 (±2.983)	10.105 (±0.148)	8.315 (±0.187)
	75	**21.567 (±4.337)**	**17.220 (±3.483)**	**0.967 (±2.883)**	**3.333 (±1.918)**	**10.943 (±0.117)**	**6.805 (±0.243)**
	100	18.133 (±12.074)	14.507 (±9.659)	1.033 (±3.057)	7.667 (±4.436)	11.049 (±0.117)	5.666 (±0.476)
74	25	19.856 (±5.372)	11.556 (±3.546)	1.067 (±2.677)	5.100 (±3.356)	8.773 (±0.140)	9.587 (±0.098)
	50	18.422 (±4.239)	11.989 (±2.828)	1.300 (±3.564)	4.533 (±2.161)	10.090 (±0.201)	8.407 (±0.168)
	75	19.900 (±4.957)	13.267 (±3.305)	0.967 (±2.883)	4.467 (±2.240)	10.962 (±0.147)	6.799 (±0.283)
	100	21.600 (±10.981)	14.400 (±7.321)	0.867 (±2.161)	5.867 (±3.674)	11.050 (±0.146)	5.877 (±0.313)
112	25	20.267 (±5.717)	9.250 (±3.036)	1.133 (±2.474)	5.167 (±3.842)	8.547 (±0.250)	9.708 (±0.042)
	50	16.867 (±3.721)	8.367 (±1.934)	0.967 (±2.512)	4.300 (±2.215)	10.073 (±0.165)	8.364 (±0.172)
	75	20.200 (±5.006)	10.100 (±2.503)	1.133 (±2.849)	4.000 (±1.965)	10.947 (±0.133)	6.850 (±0.229)
	100	22.700 (±11.870)	11.350 (±5.935)	1.033 (±3.243)	7.033 (±6.457)	11.077 (±0.129)	5.587 (±0.442)
150	25	20.133 (±5.387)	0.000 (±0.000)	0.967 (±2.327)	5.200 (±3.428)	8.603 (±0.197)	9.708 (±0.041)
	50	16.367 (±3.222)	0.000 (±0.000)	1.067 (±2.490)	4.133 (±2.193)	10.028 (±0.164)	8.440 (±0.134)
	75	19.867 (±6.252)	0.000 (±0.000)	1.067 (±2.864)	3.600 (±2.044)	10.936 (±0.122)	6.859 (±0.252)
	100	18.200 (±9.956)	0.000 (±0.000)	0.933 (±3.073)	7.400 (±6.268)	11.074 (±0.155)	5.628 (±0.501)

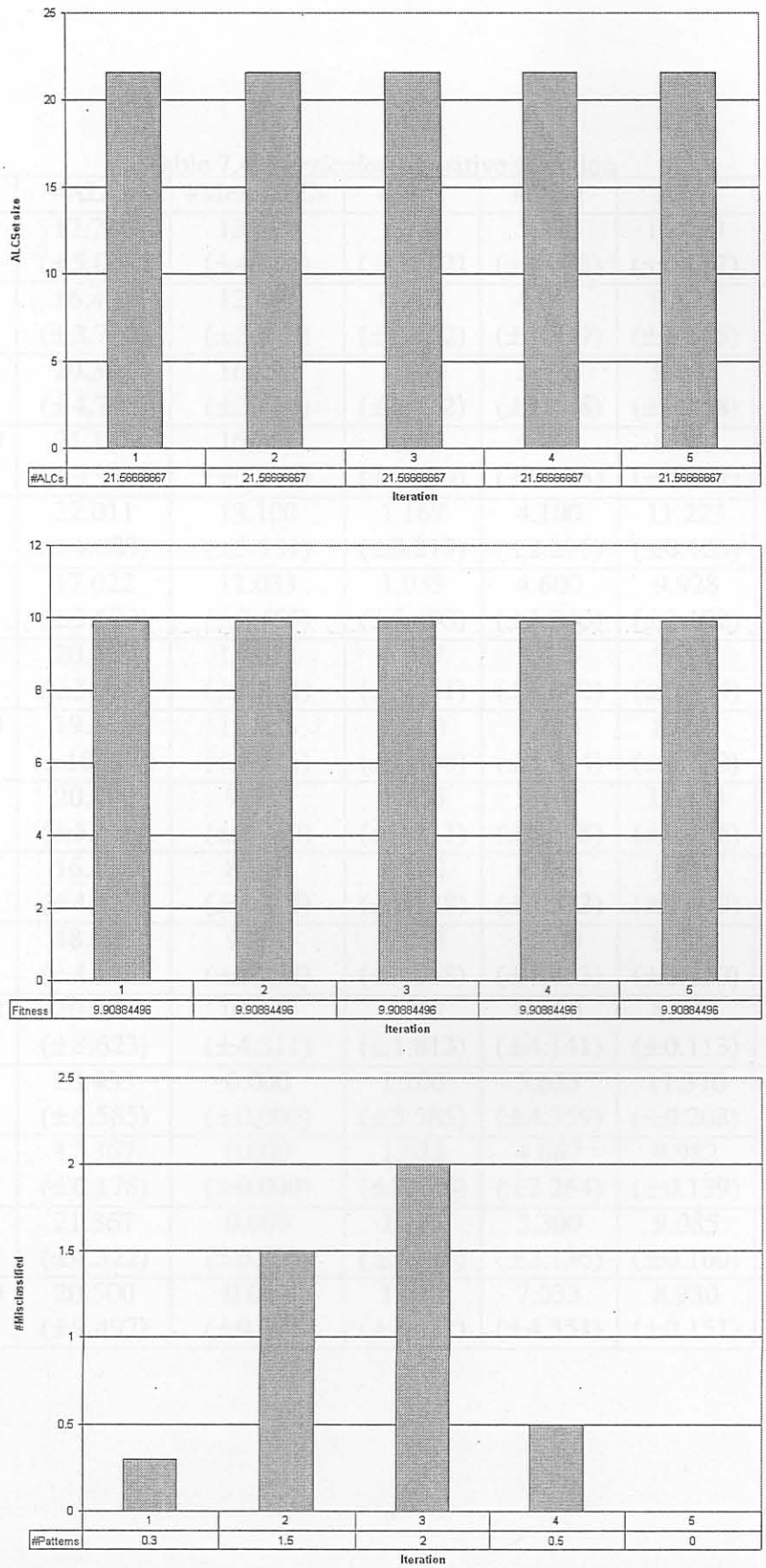


Figure 7.3: *Versicolor* - Negative selection with IS=37 and W1=75

Table 7.4: *Versicolor* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	17.780 (±5.001)	12.160 (±4.066)	1.133 (±2.662)	5.300 (±4.078)	11.039 (±0.177)	9.348 (±0.330)
	50	16.473 (±3.780)	12.607 (±3.018)	0.967 (±2.512)	4.067 (±1.999)	9.875 (±0.165)	8.277 (±0.226)
	75	**20.333 (±4.700)**	**16.253 (±3.750)**	**1.167 (±2.842)**	**3.333 (±1.768)**	**9.041 (±0.118)**	**6.757 (±0.256)**
	100	21.133 (±9.916)	16.887 (±7.942)	1.033 (±3.429)	6.067 (±4.025)	8.919 (±0.137)	5.784 (±0.400)
74	25	22.011 (±4.683)	13.100 (±3.131)	1.167 (±3.217)	4.100 (±2.295)	11.223 (±0.125)	9.624 (±0.083)
	50	17.022 (±3.529)	11.033 (±2.455)	1.033 (±2.498)	4.600 (±1.940)	9.928 (±0.192)	8.410 (±0.173)
	75	20.433 (±3.636)	13.622 (±2.424)	1.067 (±3.051)	3.833 (±1.642)	9.070 (±0.134)	6.802 (±0.257)
	100	19.133 (±10.224)	12.756 (±6.816)	0.900 (±2.708)	7.233 (±5.104)	8.951 (±0.123)	5.601 (±0.488)
112	25	20.700 (±5.676)	9.483 (±2.740)	1.133 (±3.411)	5.467 (±5.178)	11.404 (±0.196)	9.710 (±0.038)
	50	16.433 (±4.523)	8.050 (±2.175)	0.900 (±1.788)	4.433 (±2.373)	9.946 (±0.179)	8.417 (±0.205)
	75	18.600 (±4.515)	9.300 (±2.258)	0.967 (±3.068)	3.800 (±1.883)	9.022 (±0.150)	6.743 (±0.283)
	100	20.833 (±8.623)	10.417 (±4.311)	0.767 (±1.813)	6.600 (±4.141)	8.945 (±0.113)	5.823 (±0.377)
150	25	20.433 (±6.585)	0.000 (±0.000)	1.200 (±3.585)	5.633 (±4.359)	11.376 (±0.208)	9.685 (±0.049)
	50	17.367 (±6.178)	0.000 (±0.000)	1.033 (±2.684)	4.667 (±2.264)	9.982 (±0.139)	8.429 (±0.144)
	75	21.367 (±4.522)	0.000 (±0.000)	1.267 (±3.383)	3.300 (±2.136)	9.085 (±0.160)	6.894 (±0.284)
	100	20.500 (±9.497)	0.000 (±0.000)	1.033 (±3.057)	7.033 (±4.351)	8.930 (±0.151)	5.784 (±0.370)

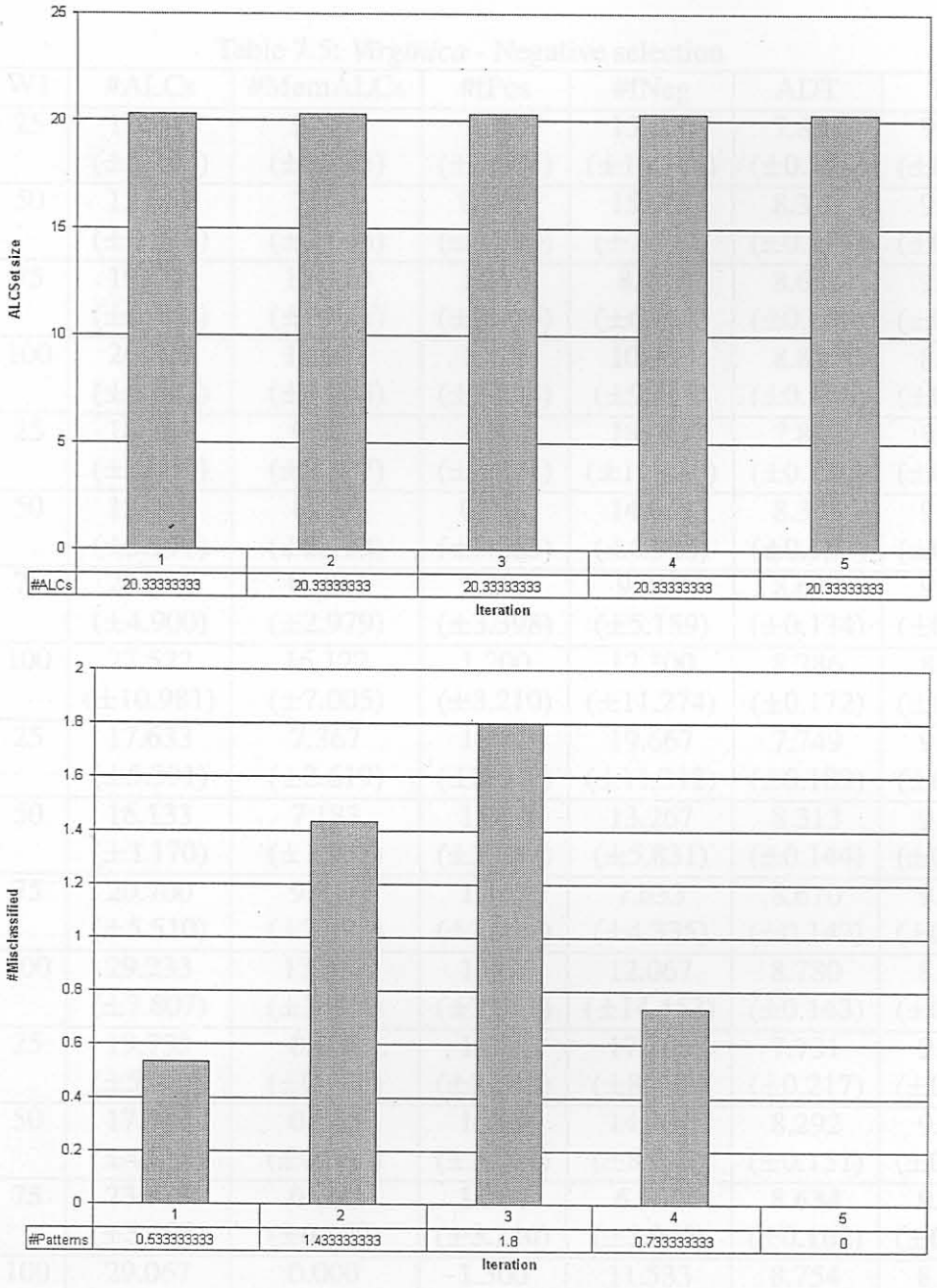


Figure 7.4: *Versicolor* - Positive selection with IS=37 and W1=75

Table 7.5: *Virginica* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	17.293 (±5.395)	9.873 (±3.835)	1.067 (±3.051)	15.000 (±10.342)	7.859 (±0.155)	9.604 (±0.249)
	50	13.667 (±3.779)	7.973 (±2.565)	0.967 (±2.883)	15.967 (±7.327)	8.327 (±0.189)	9.271 (±0.214)
	75	19.420 (±6.566)	12.920 (±4.881)	1.300 (±3.186)	8.467 (±6.962)	8.688 (±0.144)	9.020 (±0.278)
	100	26.527 (±6.502)	17.993 (±4.725)	1.133 (±3.224)	10.800 (±9.845)	8.814 (±0.156)	8.055 (±0.414)
74	25	18.289 (±4.930)	9.222 (±2.807)	1.100 (±3.231)	18.067 (±10.110)	7.812 (±0.150)	9.823 (±0.068)
	50	15.767 (±3.177)	8.356 (±2.192)	0.967 (±3.253)	14.933 (±6.913)	8.323 (±0.137)	9.466 (±0.117)
	75	20.811 (±4.900)	11.911 (±2.979)	1.200 (±3.398)	9.267 (±5.159)	8.672 (±0.134)	9.082 (±0.250)
	100	27.522 (±10.981)	16.122 (±7.005)	1.200 (±3.210)	12.300 (±11.274)	8.786 (±0.172)	8.059 (±0.383)
112	25	17.633 (±5.391)	7.367 (±2.619)	1.000 (±2.505)	19.667 (±11.312)	7.749 (±0.182)	9.881 (±0.038)
	50	16.133 (±3.170)	7.183 (±1.517)	1.067 (±3.237)	13.267 (±5.831)	8.313 (±0.144)	9.559 (±0.088)
	75	20.700 (±5.510)	9.817 (±2.490)	1.167 (±3.217)	7.633 (±4.335)	8.670 (±0.142)	9.155 (±0.158)
	100	29.233 (±7.807)	13.333 (±3.724)	1.067 (±3.051)	12.067 (±14.453)	8.780 (±0.143)	8.079 (±0.428)
150	25	19.733 (±5.401)	0.000 (±0.000)	1.033 (±3.243)	17.200 (±8.438)	7.731 (±0.217)	9.896 (±0.029)
	50	17.333 (±4.054)	0.000 (±0.000)	1.100 (±3.044)	14.067 (±8.952)	8.292 (±0.151)	9.564 (±0.065)
	75	**23.533 (±5.198)**	**0.000 (±0.000)**	**1.200 (±3.210)**	**6.900 (±3.942)**	**8.634 (±0.162)**	**9.179 (±0.197)**
	100	29.067 (±7.061)	0.000 (±0.000)	1.300 (±3.375)	11.533 (±8.525)	8.754 (±0.143)	8.067 (±0.713)

Table 7.6: *Virginica* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
37	25	16.280 (±4.784)	9.207 (±3.385)	1.033 (±3.057)	16.767 (±11.548)	12.128 (±0.186)	9.605 (±0.181)
	50	13.373 (±3.344)	7.887 (±2.432)	1.067 (±2.864)	14.633 (±7.531)	11.620 (±0.163)	9.273 (±0.170)
	75	20.453 (±6.202)	13.367 (±4.763)	1.033 (±3.057)	9.767 (±7.890)	11.342 (±0.129)	9.113 (±0.207)
	100	24.087 (±7.786)	15.860 (±5.445)	1.033 (±3.057)	13.967 (±12.505)	11.207 (±0.168)	7.976 (±0.695)
74	25	18.011 (±5.426)	8.711 (±2.904)	0.900 (±2.893)	17.533 (±11.116)	12.192 (±0.157)	9.799 (±0.090)
	50	16.622 (±3.675)	9.056 (±2.370)	1.067 (±3.051)	14.600 (±7.166)	11.681 (±0.144)	9.520 (±0.092)
	75	20.178 (±5.535)	11.367 (±3.187)	1.100 (±2.857)	8.833 (±5.565)	11.375 (±0.143)	9.150 (±0.180)
	100	26.578 (±9.091)	15.211 (±5.456)	1.167 (±3.405)	14.033 (±13.265)	11.175 (±0.141)	7.933 (±0.525)
112	25	19.733 (±5.626)	8.267 (±2.605)	1.200 (±3.210)	16.833 (±11.946)	12.264 (±0.183)	9.890 (±0.038)
	50	15.633 (±3.168)	6.850 (±1.549)	1.167 (±3.030)	14.833 (±7.235)	11.695 (±0.169)	9.529 (±0.102)
	75	**23.833 (±4.793)**	**11.083 (±2.271)**	**1.267 (±3.194)**	**6.800 (±4.552)**	**11.342 (±0.127)**	**9.171 (±0.172)**
	100	28.000 (±8.208)	13.033 (±3.859)	1.167 (±3.030)	11.033 (±7.073)	11.234 (±0.165)	8.129 (±0.309)
150	25	21.700 (±5.292)	0.000 (±0.000)	1.100 (±2.857)	13.133 (±10.884)	12.195 (±0.158)	9.895 (±0.025)
	50	15.900 (±3.836)	0.000 (±0.000)	1.167 (±3.217)	14.367 (±5.295)	11.702 (±0.193)	9.557 (±0.093)
	75	20.400 (±5.184)	0.000 (±0.000)	1.167 (±3.030)	7.967 (±5.449)	11.337 (±0.167)	9.157 (±0.163)
	100	28.933 (±9.161)	0.000 (±0.000)	1.233 (±3.202)	11.500 (±10.875)	11.221 (±0.149)	8.105 (±0.325)

Table 7.6 shows that the best classification is achieved with IS=112 and W1=75 when training the ALCs with positive selection. These parameter settings gave the lowest misclassification of 8.067 patterns ($\#Misclassified = 1.267 + 6.800 = 8.067$) with an average number of 23.833 ALCs in the active set of ALCs that had an average of 11.083 ALCs with memory status. This gives a correct classification rate of 94.622%. The size of the active set of ALCs was constant on an average of 23.833 over time. The size of the active set of ALCs was constant on 23.833 over all iterations and the average fitness of the ALC set was constant.

7.1.4 Conclusion: Iris

With *setosa* or *versicolor* as the self class, training the ALCs with the negative selection method resulted in better classification than training with the positive selection method. When patterns of the *virginica* class was used as the self set the ALCs trained with positive selection had better classification than the ALCs trained with negative selection. There is also a decrease or constant average number of ALCs in the active set of ALCs for both negative and positive selection methods with the different classes as self. The decrease or constant average number of ALCs in the set results in the decrease or constant average fitness of the ALC set.

7.2 Wisconsin Breast Cancer

The Wisconsin breast cancer data set consists of 699 patterns that are distributed between 2 classes, namely benign and malignant. Each pattern consists of 9 attributes with values in the range [1,10]. The tenth attribute is the pattern's sample code number and uniquely identifies the pattern in the data set. The sample code number was left out in the training and testing of the AIS model. There are 16 missing attribute values for the bare nuclei attribute in the data set. 458 patterns are of the benign class and 241 patterns of the malignant class. The patterns were converted to binary strings of length 36. The missing values were represented by binary strings as straight 1's.

7.2.1 Benign

Table 7.7 shows the results for classifying the Wisconsin breast cancer data set with patterns of the benign class as the self set. The ALCs were trained with the negative selection method. The best classification result was achieved when IS=524 and W1=25. An average number of 34.567 ALCs formed part of the active set of ALCs. An average number of 16.867 of the ALCs in the

Table 7.7: *Benign* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
175	25	32.492 (±5.587)	22.367 (±4.342)	1.333 (±0.922)	7.400 (±7.859)	16.159 (±0.154)	17.417 (±0.102)
	50	21.858 (±5.735)	16.058 (±4.415)	1.133 (±0.973)	14.900 (±7.685)	18.052 (±0.195)	15.659 (±0.268)
	75	29.400 (±7.342)	22.033 (±5.521)	1.067 (±1.048)	9.633 (±3.746)	19.377 (±0.135)	13.516 (±0.254)
	100	37.633 (±12.936)	28.217 (±9.724)	1.200 (±1.031)	12.267 (±7.220)	19.764 (±0.121)	12.226 (±0.328)
350	25	33.400 (±5.190)	16.300 (±2.531)	1.233 (±1.006)	7.033 (±4.605)	16.060 (±0.134)	17.503 (±0.036)
	50	21.367 (±5.385)	10.583 (±2.678)	0.767 (±0.774)	13.167 (±6.422)	17.989 (±0.183)	15.786 (±0.186)
	75	27.500 (±7.960)	13.750 (±3.980)	0.933 (±0.740)	11.567 (±6.021)	19.384 (±0.171)	13.516 (±0.342)
	100	36.100 (±11.678)	18.050 (±5.839)	1.233 (±1.040)	11.267 (±7.012)	19.779 (±0.132)	12.271 (±0.492)
524	**25**	**34.567 (±5.029)**	**16.867 (±2.526)**	**1.367 (±1.189)**	**6.267 (±5.219)**	**16.061 (±0.188)**	**17.499 (±0.053)**
	50	20.967 (±5.021)	10.400 (±2.558)	1.000 (±0.871)	13.667 (±5.785)	17.963 (±0.163)	15.773 (±0.142)
	75	29.167 (±8.550)	14.583 (±4.275)	1.033 (±0.890)	9.933 (±5.675)	19.390 (±0.158)	13.508 (±0.270)
	100	39.167 (±11.908)	19.583 (±5.954)	0.933 (±0.785)	10.367 (±8.109)	19.795 (±0.156)	12.204 (±0.598)
699	25	33.100 (±6.467)	0.000 (±0.000)	1.400 (±0.932)	8.067 (±10.945)	16.023 (±0.233)	17.506 (±0.050)
	50	20.867 (±5.419)	0.000 (±0.000)	1.200 (±1.031)	12.433 (±5.643)	17.991 (±0.203)	15.764 (±0.211)
	75	27.933 (±8.440)	0.000 (±0.000)	0.933 (±0.868)	10.567 (±4.150)	19.390 (±0.136)	13.503 (±0.233)
	100	37.467 (±13.405)	0.000 (±0.000)	1.200 (±1.031)	11.700 (±7.274)	19.773 (±0.136)	12.193 (±0.443)

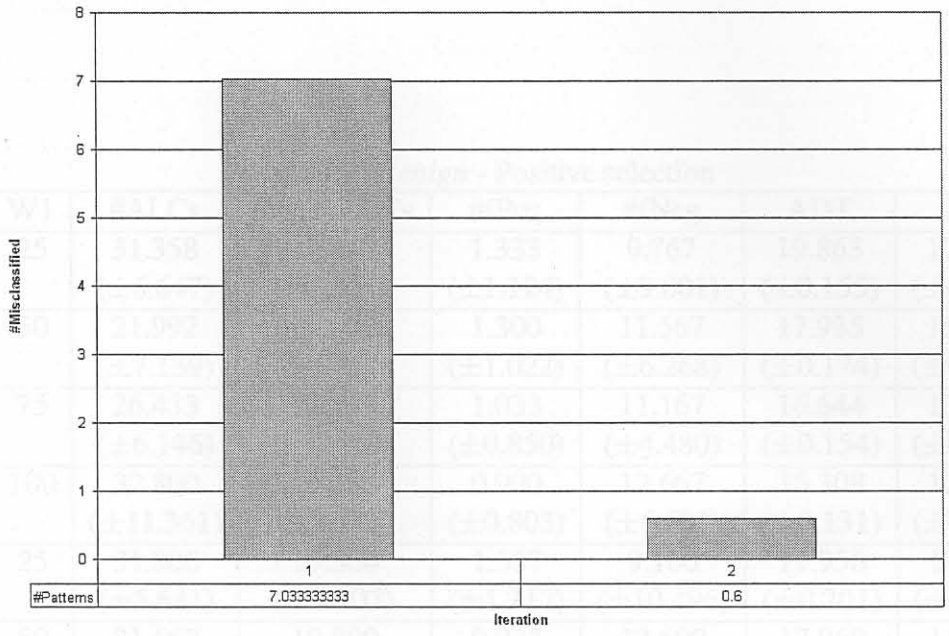


Figure 7.5: *Benign* - Negative selection with IS=524 and W1=25

active set had memory status. The ALCs classified 6.267 patterns falsely as benign and 1.367 as malignant. This gives a misclassification of 7.634 patterns ($\#Misclassified = 6.267 + 1.367 = 7.634$) and a correct classification of 98.907%. The size of the active set of ALCs was constant on 34.567 over all iterations and thus the average fitness of the ALC set was constant. Figure 7.5 shows a decrease in the average number of misclassification over the iterations.

Table 7.8 shows the results when the ALCs were trained with the positive selection method. The best classification result was achieved when IS=524 and W1=25. An average number of 33.800 ALCs formed part of the active set of ALCs. An average number of 16.467 of the ALCs in the active set had memory status. The ALCs classified 6.900 patterns falsely as benign and 1.500 as malignant which gave a misclassification of 8.400 patterns ($\#Misclassified = 6.900 + 1.500 = 8.400$) and a correct classification rate of 98.798%. The number of ALCs in the active set of ALCs was constant on 33.800 over all iterations with a constant average fitness.

7.2.2 Malignant

The results for training the ALCs with the negative selection method is shown in table 7.9. The patterns of the malignant class was used as the self set. The best classification result was

Table 7.8: *Benign* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
175	25	31.358 (±6.647)	21.442 (±5.110)	1.333 (±1.124)	9.767 (±9.001)	19.863 (±0.155)	17.401 (±0.089)
	50	21.992 (±7.139)	16.092 (±5.411)	1.300 (±1.022)	11.567 (±6.268)	17.935 (±0.174)	15.701 (±0.204)
	75	26.433 (±6.146)	19.792 (±4.614)	1.033 (±0.850)	11.167 (±4.480)	16.644 (±0.154)	13.530 (±0.324)
	100	32.800 (±11.361)	24.583 (±8.516)	0.900 (±0.803)	12.667 (±6.504)	16.308 (±0.131)	12.256 (±0.250)
350	25	31.900 (±5.641)	15.500 (±2.907)	1.367 (±1.217)	9.100 (±10.496)	19.956 (±0.201)	17.485 (±0.054)
	50	21.667 (±7.448)	10.800 (±3.732)	0.933 (±0.868)	13.600 (±8.046)	17.960 (±0.180)	15.694 (±0.185)
	75	28.600 (±6.667)	14.300 (±3.334)	1.267 (±0.907)	10.267 (±3.956)	16.622 (±0.145)	13.507 (±0.249)
	100	34.967 (±12.524)	17.483 (±6.262)	0.700 (±0.750)	11.867 (±6.942)	16.245 (±0.141)	12.262 (±0.321)
524	**25**	**33.800 (±3.398)**	**16.467 (±1.756)**	**1.500 (±1.075)**	**6.900 (±3.284)**	**19.922 (±0.117)**	**17.490 (±0.044)**
	50	21.867 (±5.619)	10.917 (±2.801)	1.133 (±0.937)	13.833 (±7.027)	18.001 (±0.149)	15.755 (±0.164)
	75	29.367 (±7.388)	14.683 (±3.694)	1.067 (±0.907)	10.000 (±5.675)	16.619 (±0.140)	13.429 (±0.312)
	100	38.133 (±12.875)	19.067 (±6.438)	1.200 (±1.031)	11.767 (±10.061)	16.214 (±0.207)	12.168 (±0.424)
699	25	32.900 (±6.989)	0.000 (±0.000)	1.600 (±1.404)	9.733 (±14.064)	19.958 (±0.258)	17.477 (±0.044)
	50	21.400 (±5.302)	0.000 (±0.000)	1.100 (±0.995)	11.467 (±5.710)	17.964 (±0.153)	15.710 (±0.183)
	75	27.200 (±7.179)	0.000 (±0.000)	1.000 (±0.983)	10.467 (±4.439)	16.637 (±0.175)	13.512 (±0.353)
	100	38.433 (±14.943)	0.000 (±0.000)	1.100 (±0.960)	11.500 (±8.228)	16.234 (±0.144)	12.125 (±0.451)

Table 7.9: *Malignant* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
175	25	23.342 (±3.960)	6.758 (±1.864)	2.633 (±1.189)	146.100 (±98.744)	13.923 (±0.162)	16.963 (±0.355)
	50	15.108 (±3.566)	5.650 (±1.666)	1.867 (±1.074)	98.533 (±69.563)	15.024 (±0.158)	15.835 (±0.420)
	75	19.133 (±4.832)	9.733 (±2.683)	1.833 (±1.367)	46.333 (±47.790)	15.954 (±0.138)	14.249 (±0.318)
	100	25.875 (±9.360)	13.975 (±5.936)	1.533 (±1.042)	62.200 (±71.823)	16.276 (±0.132)	12.756 (±0.663)
350	25	31.000 (±5.855)	7.867 (±2.432)	2.067 (±1.437)	117.633 (±95.248)	13.530 (±0.217)	17.739 (±0.038)
	50	21.033 (±6.856)	7.300 (±2.507)	1.967 (±1.326)	115.100 (±83.162)	14.842 (±0.159)	16.663 (±0.144)
	75	23.200 (±5.209)	9.500 (±2.338)	1.733 (±0.980)	46.500 (±24.221)	15.910 (±0.114)	14.664 (±0.234)
	100	34.900 (±13.510)	14.067 (±5.889)	1.700 (±1.393)	52.200 (±52.346)	16.271 (±0.139)	13.010 (±0.397)
524	25	30.800 (±6.885)	8.683 (±2.329)	2.300 (±1.418)	149.900 (±89.560)	13.563 (±0.165)	17.735 (±0.037)
	50	20.667 (±5.726)	7.750 (±2.586)	1.800 (±1.297)	108.633 (±89.016)	14.799 (±0.169)	16.710 (±0.188)
	75	25.133 (±6.709)	10.917 (±3.235)	1.967 (±1.129)	60.367 (±48.565)	15.918 (±0.130)	14.620 (±0.336)
	100	32.300 (±13.018)	14.133 (±5.810)	1.633 (±1.273)	49.833 (±54.484)	16.255 (±0.137)	13.002 (±0.463)
699	25	32.433 (±5.137)	0.000 (±0.000)	2.567 (±1.382)	112.500 (±68.027)	13.597 (±0.134)	17.732 (±0.037)
	50	20.500 (±5.036)	0.000 (±0.000)	1.933 (±1.337)	90.533 (±66.951)	14.848 (±0.122)	16.685 (±0.163)
	75	24.133 (±6.163)	0.000 (±0.000)	1.800 (±1.243)	45.133 (±25.962)	15.927 (±0.126)	14.613 (±0.295)
	100	**33.033 (±10.791)**	**0.000 (±0.000)**	**1.767 (±1.251)**	**44.400 (±36.541)**	**16.267 (±0.136)**	**13.041 (±0.587)**

achieved when IS=699 and W1=100 which gave an average number of 33.033 ALCs in the active set of ALCs. An average number of 0.000 of the ALCs in the active set have memory status since there is only one iteration. The ALCs classified 44.400 patterns falsely as malignant and 1.767 as benign giving a misclassification of 46.167 patterns ($\#Misclassified = 44.400 + 1.767 = 46.167$) which gives a correct classification rate of 93.395%.

Table 7.10 shows the results when the ALCs were trained with the positive selection method. The best classification result was achieved when IS=175 and W1=75 which gave an average number of 22.308 ALCs in the active set of ALCs. An average number of 11.525 of the ALCs in the active set had memory status. The ALCs classified 40.333 patterns falsely as malignant and 1.967 as benign which gives a misclassification of 42.300 patterns ($\#Misclassified = 40.333 + 1.967 = 42.300$) and a correct classification rate of 93.948%. Figure 7.5 shows that the initial number of ALCs in the active set of ALCs was on average 26.566 and decreased over four iterations to an average of 15.333. Therefor the average fitness of the ALC set decreased over all iterations from 15.588 to 15.458. The average number of misclassification increased from iteration one to two and then decreased to iteration four.

7.2.3 Conclusion: Wisconsin Breast Cancer

In conclusion, the ALCs trained on the *benign* class with negative selection had a correct classification of 98.907% with an average number of 34.567 ALCs that had an average HD of 17.499. The ALCs that had been trained with positive selection on the *benign* class as self had a correct classification of 98.798% with an average number of 33.800 ALCs that had an average HD of 17.490. Comparing these results shows that there is neither a major difference in correct classification between the two different training methods nor in the average number of ALCs or HD with the same parameter settings (IS=524, W1=25), though negative selection does have a slightly better correct classification than positive selection. With the patterns from the *malignant* class as the self set the negative selection method had a correct classification of 93.395% with an average number of 33.033 ALCs that had an average HD of 13.041. The positive selection method had a correct classification of 93.948% with an average number of 22.308 ALCs that had an average HD of 14.322. These results conclude that when the patterns of the *malignant* class is used as the self set, different parameter settings are necessary to achieve similar correct classification results for both the selection methods and that the average number of ALCs for negative selection is higher than the average number of ALCs for positive selection. The difference in the average number of ALCs indicates that the patterns from the *benign* class have a

Table 7.10: *Malignant* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
175	25	22.333 (±4.240)	6.350 (±1.637)	2.767 (±1.695)	126.000 (±90.812)	22.037 (±0.151)	16.800 (±0.360)
	50	16.275 (±5.096)	6.375 (±2.677)	1.367 (±1.217)	106.167 (±76.795)	21.053 (±0.181)	15.923 (±0.420)
	75	**22.308 (±5.555)**	**11.525 (±3.397)**	**1.967 (±1.377)**	**40.333 (±19.752)**	**20.043 (±0.121)**	**14.322 (±0.267)**
	100	26.117 (±8.666)	13.933 (±5.283)	1.467 (±1.456)	62.733 (±75.871)	19.680 (±0.129)	12.643 (±0.633)
350	25	31.833 (±6.701)	7.950 (±2.291)	2.000 (±1.259)	126.500 (±99.243)	22.419 (±0.100)	17.741 (±0.043)
	50	21.100 (±4.708)	6.967 (±1.857)	2.000 (±1.390)	98.167 (±60.796)	21.190 (±0.136)	16.696 (±0.169)
	75	24.667 (±6.315)	9.733 (±2.693)	1.967 (±1.564)	60.800 (±47.323)	20.111 (±0.109)	14.685 (±0.309)
	100	28.867 (±13.577)	11.817 (±5.759)	2.067 (±1.530)	61.400 (±50.348)	19.798 (±0.169)	13.080 (±0.439)
524	25	32.500 (±4.524)	8.933 (±1.973)	1.933 (±1.437)	109.667 (±61.938)	22.407 (±0.097)	17.741 (±0.034)
	50	20.400 (±6.414)	7.300 (±2.753)	2.033 (±1.217)	115.767 (±86.717)	21.225 (±0.177)	16.698 (±0.148)
	75	25.700 (±6.944)	11.167 (±3.049)	1.967 (±0.928)	54.600 (±41.608)	20.026 (±0.104)	14.581 (±0.303)
	100	32.267 (±14.350)	14.200 (±6.504)	1.933 (±1.437)	55.100 (±61.477)	19.752 (±0.140)	12.931 (±0.530)
699	25	28.600 (±9.227)	0.000 (±0.000)	1.867 (±1.408)	164.033 (±105.704)	22.505 (±0.209)	17.723 (±0.049)
	50	19.800 (±5.054)	0.000 (±0.000)	2.000 (±1.114)	114.900 (±79.443)	21.182 (±0.106)	16.669 (±0.125)
	75	26.867 (±4.939)	0.000 (±0.000)	1.800 (±1.095)	45.800 (±21.815)	20.093 (±0.135)	14.733 (±0.294)
	100	29.333 (±10.908)	0.000 (±0.000)	1.667 (±1.241)	60.900 (±64.088)	19.712 (±0.143)	13.016 (±0.481)

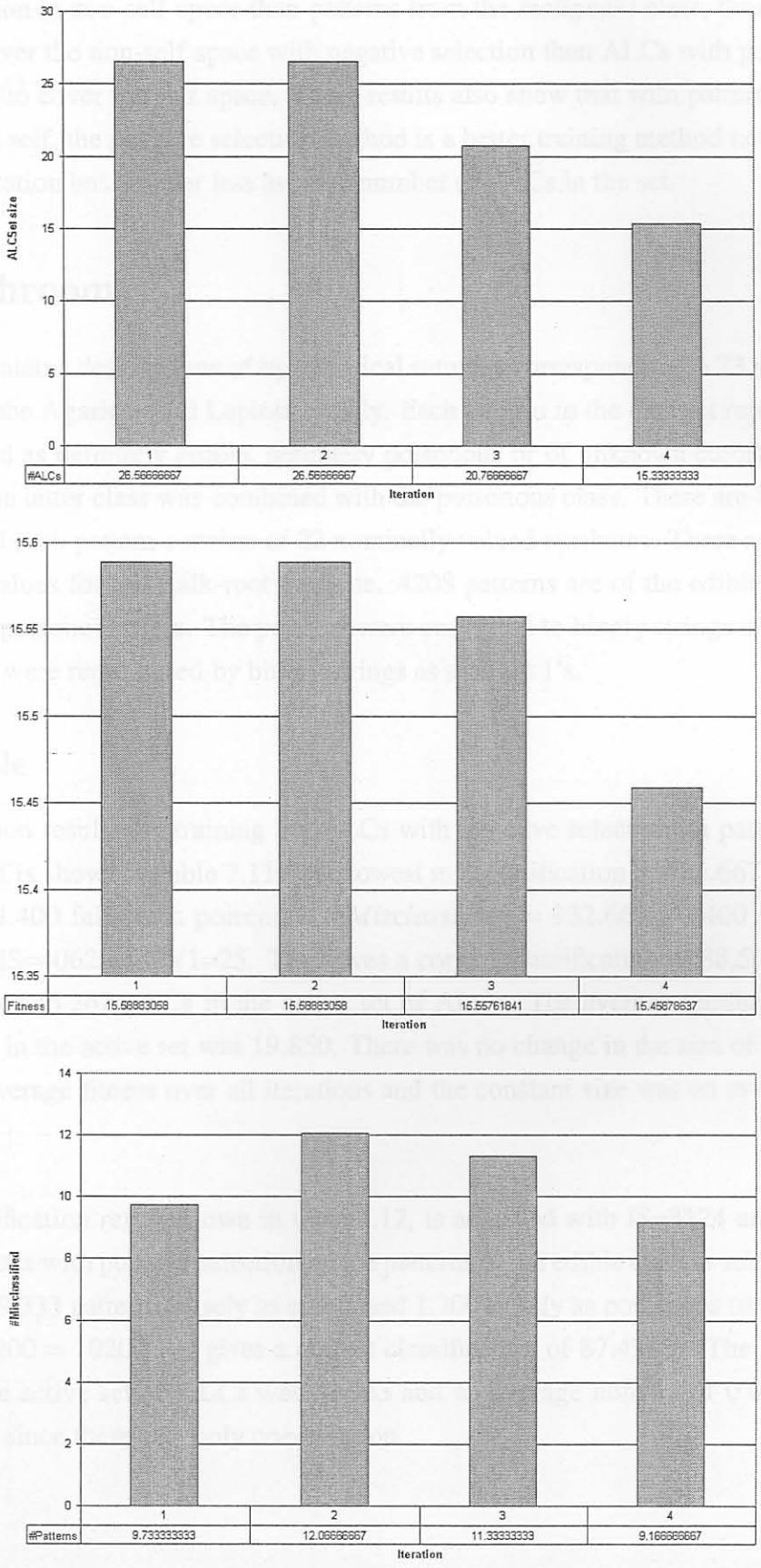


Figure 7.6: *Malignant* - Positive selection with IS=175 and W1=75

larger distribution in non-self space than patterns from the *malignant* class, thus more ALCs is necessary to cover the non-self space with negative selection than ALCs with positive selection that only needs to cover the self space. These results also show that with patterns from the *malignant* class as self, the positive selection method is a better training method not only for better correct classification but also for less average number of ALCs in the set.

7.3 Mushroom

The data set contains descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family. Each pattern in the data set represents a specie that is classified as definitely edible, definitely poisonous or of unknown edibility and not recommended. The latter class was combined with the poisonous class. There are 8124 patterns in the data set and each pattern consists of 22 nominally valued attributes. There are 2480 patterns with missing values for the stalk-root attribute. 4208 patterns are of the edible class and 3916 patterns of the poisonous class. The patterns were converted to binary strings of length 57. The missing values were represented by binary strings as straight 1's.

7.3.1 Edible

The classification results for training the ALCs with negative selection on patterns of the edible class as self is shown in table 7.11. The lowest misclassification of 932.667 patterns falsely as edible and 1.400 falsely as poisonous ($\#Misclassified = 932.667 + 1.400 = 934.067$) was achieved with IS=4062 and W1=25. This gives a correct classification of 88.502% with an average number of 46.267 ALCs in the active set of ALCs. The average number of ALCs with memory status in the active set was 19.850. There was no change in the size of the active set of ALCs or the average fitness over all iterations and the constant size was an average number of 46.267 ALCs.

The best classification result shown in table 7.12, is achieved with IS=8124 and W1=25 when training the ALCs with positive selection on the patterns of the edible class as self. The misclassification of 1019.333 patterns falsely as edible and 1.200 falsely as poisonous ($\#Misclassified = 1019.333 + 1.200 = 1020.533$) gives a correct classification of 87.438%. The average number of ALCs in the active set of ALCs was 43.933 and an average number of 0.000 of these had memory status since there was only one iteration.

Table 7.11: *Edible* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
2031	25	37.925 (±4.890)	17.675 (±3.150)	0.967 (±1.351)	949.867 (±195.539)	25.682 (±0.183)	27.475 (±0.100)
	50	26.333 (±8.039)	13.625 (±4.860)	1.100 (±2.023)	1376.800 (±448.271)	28.627 (±0.148)	24.874 (±0.252)
	75	34.125 (±11.146)	19.583 (±7.072)	1.267 (±2.180)	1276.367 (±274.065)	30.695 (±0.232)	21.749 (±0.330)
	100	63.800 (±18.449)	39.900 (±12.374)	1.867 (±2.738)	1172.033 (±290.378)	31.466 (±0.193)	20.162 (±0.337)
4062	**25**	**46.267 (±3.999)**	**19.850 (±2.297)**	**1.400 (±2.222)**	**932.667 (±196.295)**	**25.520 (±0.131)**	**27.698 (±0.063)**
	50	28.067 (±8.358)	13.017 (±4.128)	1.167 (±1.704)	1451.367 (±449.082)	28.554 (±0.239)	25.009 (±0.222)
	75	43.833 (±11.641)	20.783 (±5.640)	1.700 (±2.996)	1224.100 (±359.480)	30.706 (±0.160)	21.845 (±0.361)
	100	74.733 (±18.221)	35.467 (±8.684)	2.767 (±4.368)	1083.733 (±186.907)	31.449 (±0.144)	20.224 (±0.251)
6093	25	45.033 (±4.390)	19.350 (±2.327)	2.000 (±3.373)	1021.633 (±216.756)	25.510 (±0.161)	27.704 (±0.051)
	50	26.767 (±8.097)	12.667 (±4.022)	1.067 (±1.552)	1300.867 (±314.306)	28.534 (±0.214)	25.041 (±0.204)
	75	45.800 (±9.568)	21.933 (±4.686)	1.467 (±2.417)	1221.533 (±253.926)	30.687 (±0.136)	21.879 (±0.249)
	100	63.733 (±17.416)	30.500 (±8.270)	2.667 (±3.994)	1171.700 (±202.250)	31.391 (±0.137)	20.303 (±0.296)
8124	25	45.333 (±4.011)	0.000 (±0.000)	1.533 (±1.943)	981.233 (±314.111)	25.498 (±0.154)	27.709 (±0.064)
	50	29.267 (±9.836)	0.000 (±0.000)	1.267 (±1.856)	1387.433 (±541.728)	28.609 (±0.276)	24.960 (±0.251)
	75	42.733 (±10.544)	0.000 (±0.000)	1.900 (±2.987)	1244.667 (±262.081)	30.701 (±0.213)	21.909 (±0.369)
	100	65.867 (±17.702)	0.000 (±0.000)	1.967 (±3.023)	1120.867 (±187.307)	31.451 (±0.132)	20.197 (±0.283)

Table 7.12: *Edible* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
2031	25	39.417 (±3.322)	18.500 (±2.310)	1.600 (±2.647)	1041.067 (±208.822)	31.348 (±0.167)	27.516 (±0.116)
	50	23.317 (±7.704)	11.758 (±4.421)	0.667 (±1.028)	1459.433 (±482.058)	28.467 (±0.209)	24.794 (±0.324)
	75	40.308 (±10.520)	23.492 (±7.175)	1.600 (±2.568)	1175.733 (±267.085)	26.284 (±0.164)	21.788 (±0.329)
	100	60.367 (±16.348)	37.025 (±11.319)	2.333 (±4.205)	1147.567 (±263.109)	25.501 (±0.204)	20.128 (±0.386)
4062	25	42.733 (±9.487)	17.717 (±4.211)	1.500 (±1.996)	1146.200 (±454.393)	31.546 (±0.217)	27.681 (±0.082)
	50	27.300 (±8.408)	12.633 (±3.792)	1.367 (±2.356)	1470.900 (±403.862)	28.449 (±0.218)	25.036 (±0.246)
	75	46.833 (±11.293)	22.333 (±5.284)	1.567 (±2.112)	1164.233 (±292.382)	26.275 (±0.176)	21.882 (±0.323)
	100	62.367 (±17.824)	29.683 (±8.405)	2.233 (±3.645)	1149.600 (±241.385)	25.535 (±0.169)	20.176 (±0.290)
6093	25	44.033 (±7.323)	19.450 (±3.539)	1.300 (±1.705)	1098.233 (±398.745)	31.574 (±0.260)	27.707 (±0.051)
	50	28.367 (±8.344)	13.350 (±4.067)	0.867 (±1.634)	1437.400 (±409.499)	28.416 (±0.225)	25.027 (±0.218)
	75	44.467 (±11.073)	21.367 (±5.597)	1.967 (±2.942)	1212.500 (±264.740)	26.388 (±0.214)	22.012 (±0.394)
	100	68.067 (±18.431)	32.583 (±8.747)	3.133 (±4.547)	1205.900 (±229.452)	25.643 (±0.176)	20.350 (±0.327)
8124	**25**	**43.933 (±7.506)**	**0.000 (±0.000)**	**1.200 (±1.495)**	**1019.333 (±373.054)**	**31.502 (±0.153)**	**27.690 (±0.059)**
	50	29.533 (±8.464)	0.000 (±0.000)	1.100 (±2.040)	1322.200 (±332.097)	28.340 (±0.202)	24.925 (±0.238)
	75	39.200 (±10.584)	0.000 (±0.000)	1.800 (±2.809)	1250.800 (±308.670)	26.322 (±0.187)	21.865 (±0.361)
	100	71.133 (±20.669)	0.000 (±0.000)	2.167 (±3.064)	1205.633 (±283.299)	25.535 (±0.158)	20.159 (±0.339)

7.3.2 Poisonous

The results for training the ALCs with the negative selection method on patterns of the poisonous class as self is shown in table 7.13. The best classification result was achieved when IS=8124 and W1=25 which gave an average number of 45.033 ALCs in the active set of ALCs. An average number of 0.000 of the ALCs in the active set had memory status since there is only one iteration. The ALCs classified 1591.933 patterns falsely as poisonous and 1.700 as edible giving a misclassification of 1593.633 patterns ($\#Misclassified = 1591.933 + 1.700 = 1593.633$) which gives a correct classification rate of 80.383%.

The best classification result shown in table 7.14, was achieved with IS=2031 and W1=100 when training the ALCs with positive selection on the patterns of the poisonous class as self. The misclassification of 1695.000 patterns falsely as poisonous and 2.467 falsely as edible patterns ($\#Misclassified = 1695.000 + 2.467 = 1697.467$) gives a correct classification of 79.105%. The average number of ALCs in the active set of ALCs was 50.717 and an average number of 17.242 of these had memory status. Figure 7.6 shows that the initial number of ALCs in the active set of ALCs was on average 64.033 and decreased over four iterations to an average of 26.166, but different from the previous experiments the average fitness of the ALC set increased over four iterations from 31.278 to 31.429. The average number of misclassification increased from iteration one to two and then decreased to iteration four.

7.3.3 Conclusion: Mushroom

Comparing the different selection methods with patterns from the *edible* class as the self set, it can be concluded that the negative selection method has a slightly better correct classification than the positive selection method. Both the selection methods have a value of 25 for W1 but different values for IS to obtain the best classification results. The same value of 25 for W1 indicates that the fitness of the ALCs in the GA for both selection methods is more influenced by their HD than by their ADT. The HD of both selection methods differ with 0.008 and indicates that the amount of overlap is more or less the same for both selection methods. When patterns from the *poisonous* class is used as the self set then the negative selection method has a better classification performance than the positive selection method since not only does the negative selection method have better correct classification results but also has on average less ALCs than the positive selection method to classify the patterns. These results indicate that the negative selection method has better classification performance than the positive selection method with

Table 7.13: *Poisonous* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
2031	25	34.925 (±5.412)	10.383 (±2.183)	2.100 (±2.928)	1876.300 (±529.147)	25.609 (±0.222)	27.371 (±0.191)
	50	23.858 (±5.654)	7.158 (±2.091)	1.833 (±1.949)	2239.167 (±530.783)	28.477 (±0.266)	24.864 (±0.365)
	75	33.575 (±10.097)	11.217 (±3.987)	2.033 (±3.243)	1917.000 (±458.657)	30.609 (±0.199)	21.728 (±0.389)
	100	52.792 (±14.391)	18.042 (±5.996)	3.033 (±4.687)	1833.933 (±425.009)	31.269 (±0.158)	20.342 (±0.285)
4062	25	46.333 (±2.551)	17.217 (±1.804)	1.400 (±1.754)	1834.900 (±382.288)	25.535 (±0.113)	27.722 (±0.048)
	50	28.533 (±8.245)	11.350 (±3.462)	1.200 (±1.375)	2254.333 (±527.809)	28.446 (±0.209)	25.154 (±0.184)
	75	44.300 (±11.928)	17.167 (±4.652)	1.867 (±2.460)	1825.600 (±342.097)	30.529 (±0.156)	22.074 (±0.263)
	100	66.400 (±21.888)	24.617 (±8.242)	2.833 (±3.505)	1686.100 (±474.694)	31.288 (±0.197)	20.340 (±0.378)
6093	25	44.700 (±7.498)	18.967 (±3.704)	1.767 (±1.478)	1823.233 (±527.523)	25.530 (±0.187)	27.718 (±0.059)
	50	29.933 (±9.025)	13.517 (±4.419)	1.300 (±2.020)	2236.967 (±462.290)	28.499 (±0.210)	25.094 (±0.282)
	75	40.800 (±10.610)	18.183 (±4.700)	1.433 (±1.813)	1952.600 (±474.423)	30.487 (±0.139)	22.035 (±0.308)
	100	63.867 (±17.338)	28.167 (±7.863)	2.467 (±2.956)	1746.633 (±462.588)	31.280 (±0.120)	20.370 (±0.224)
8124	**25**	**45.033 (±4.582)**	**0.000 (±0.000)**	**1.700 (±2.003)**	**1591.933 (±382.550)**	**25.523 (±0.142)**	**27.729 (±0.049)**
	50	27.567 (±8.736)	0.000 (±0.000)	1.367 (±1.608)	2218.267 (±498.915)	28.509 (±0.218)	25.052 (±0.235)
	75	42.167 (±13.099)	0.000 (±0.000)	1.533 (±1.925)	1889.133 (±391.633)	30.488 (±0.152)	22.089 (±0.289)
	100	57.633 (±24.102)	0.000 (±0.000)	1.867 (±2.460)	1940.833 (±569.112)	31.267 (±0.228)	20.275 (±0.522)

Table 7.14: *Poisonous* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
2031	25	36.433 (±3.570)	11.075 (±1.670)	2.033 (±2.580)	1748.767 (±335.773)	31.338 (±0.188)	27.385 (±0.114)
	50	25.033 (±6.828)	7.850 (±2.505)	1.567 (±1.888)	2095.600 (±464.683)	28.493 (±0.265)	24.844 (±0.284)
	75	32.792 (±9.468)	10.558 (±3.776)	1.500 (±1.796)	1989.933 (±504.108)	26.487 (±0.214)	21.843 (±0.372)
	100	**50.717 (±13.291)**	**17.242 (±5.158)**	**2.467 (±3.391)**	**1695.000 (±401.959)**	**25.665 (±0.178)**	**20.201 (±0.336)**
4062	25	43.867 (±8.025)	15.967 (±3.181)	1.700 (±2.395)	1813.733 (±587.445)	31.556 (±0.340)	27.737 (±0.067)
	50	28.333 (±9.060)	11.117 (±4.023)	0.867 (±1.306)	2281.133 (±449.601)	28.611 (±0.219)	25.177 (±0.212)
	75	41.200 (±9.477)	15.867 (±4.013)	1.867 (±2.330)	1949.633 (±311.235)	26.460 (±0.177)	22.024 (±0.287)
	100	59.900 (±17.574)	22.067 (±6.700)	2.600 (±3.058)	1773.900 (±491.493)	25.779 (±0.162)	20.377 (±0.287)
6093	25	45.400 (±6.946)	19.517 (±3.158)	1.933 (±2.067)	1712.333 (±479.901)	31.506 (±0.246)	27.727 (±0.054)
	50	27.100 (±8.763)	12.100 (±4.229)	1.233 (±2.096)	2221.867 (±541.412)	28.526 (±0.235)	25.096 (±0.252)
	75	39.667 (±11.133)	17.783 (±5.051)	1.700 (±1.784)	2069.133 (±343.100)	26.489 (±0.188)	21.993 (±0.394)
	100	68.233 (±21.837)	30.267 (±9.868)	2.467 (±3.137)	1704.767 (±562.829)	25.748 (±0.198)	20.400 (±0.382)
8124	25	45.733 (±3.741)	0.000 (±0.000)	1.967 (±2.632)	1843.533 (±334.025)	31.480 (±0.125)	27.725 (±0.049)
	50	31.700 (±7.996)	0.000 (±0.000)	1.500 (±2.662)	2111.500 (±360.474)	28.498 (±0.178)	25.163 (±0.213)
	75	39.867 (±11.365)	0.000 (±0.000)	1.533 (±2.097)	1886.267 (±496.066)	26.526 (±0.172)	22.101 (±0.386)
	100	68.100 (±18.054)	0.000 (±0.000)	3.100 (±3.595)	1729.967 (±474.410)	25.787 (±0.198)	20.445 (±0.302)

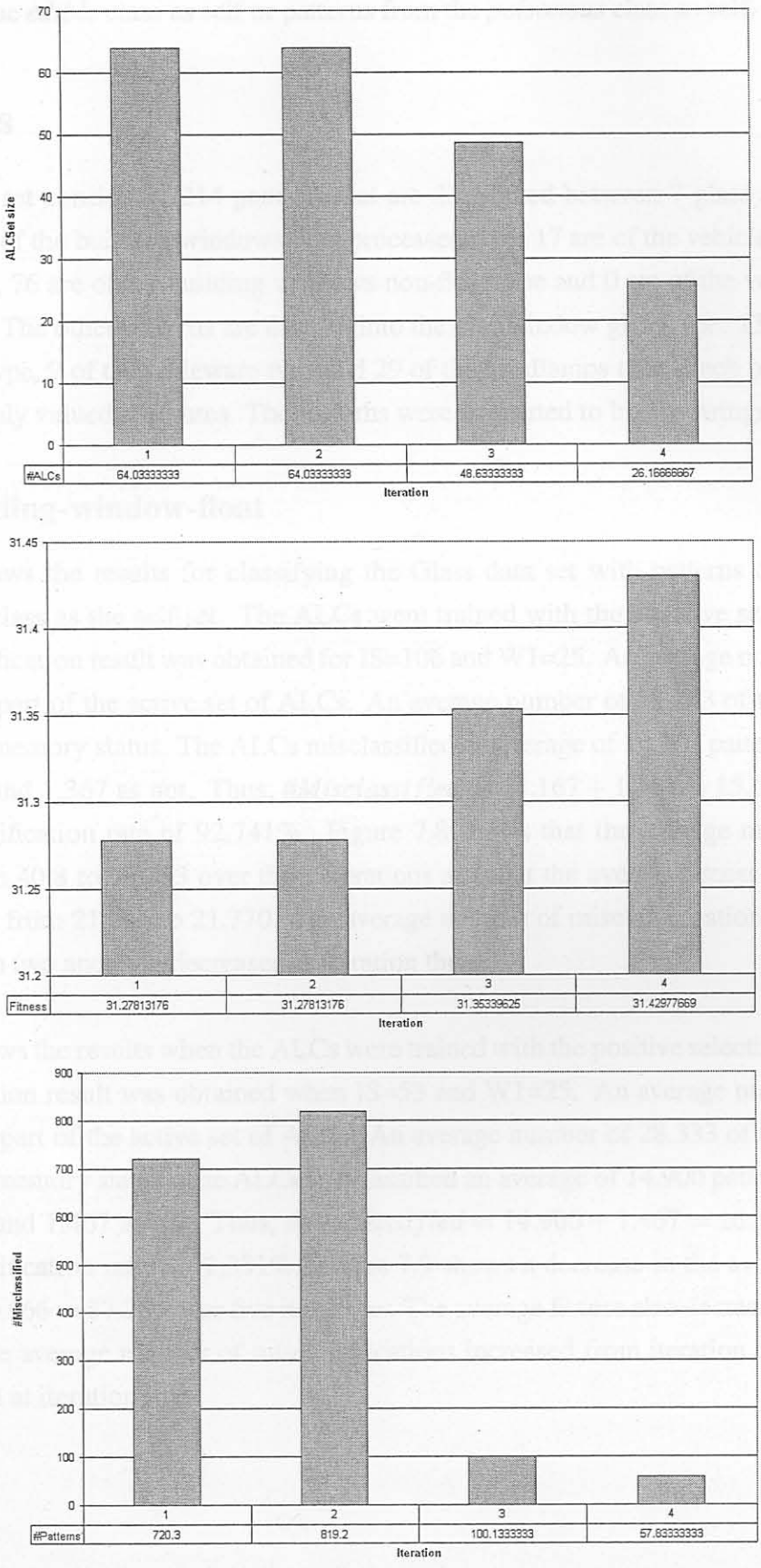


Figure 7.7: *Poisonous* - Positive selection with IS=2031 and W1=100

patterns from the *edible* class as self or patterns from the *poisonous* class as self.

7.4 Glass

The glass data set consists of 214 patterns that are distributed between 7 glass types (classes). 70 patterns are of the building windows float processed type, 17 are of the vehicle windows float processed type, 76 are of the building windows non-float type and 0 are of the vehicle windows non-float type. The other patterns are divided into the non-window glass type: 13 patterns are of the container type, 9 of the tableware type and 29 of the headlamps type. Each patterns consists of 9 continuously valued attributes. The patterns were converted to binary strings of length 45.

7.4.1 Building-window-float

Table 7.15 shows the results for classifying the Glass data set with patterns of the building-window-float class as the self set. The ALCs were trained with the negative selection method. The best classification result was obtained for IS=106 and W1=25. An average number of 40.444 ALCs formed part of the active set of ALCs. An average number of 25.233 of the ALCs in the active set had memory status. The ALCs misclassified an average of 14.167 patterns as building-window-float and 1.367 as not. Thus, $\#Misclassified = 14.167 + 1.367 = 15.534$ which gives a correct classification rate of 92.741%. Figure 7.8 shows that the average number of ALCs decreased from 40.8 to 39.733 over three iterations and that the average fitness of the ALC set also decreased from 21.796 to 21.770. The average number of misclassification increased from iteration one to two and then decreased at iteration three.

Table 7.16 shows the results when the ALCs were trained with the positive selection method. The best classification result was obtained when IS=53 and W1=25. An average number of 39.167 ALCs formed part of the active set of ALCs. An average number of 28.333 of the ALCs in the active set had memory status. The ALCs misclassified an average of 14.900 patterns as building-window-float and 1.467 as not. Thus, $\#Misclassified = 14.900 + 1.467 = 16.367$ which gives a correct classification rate of 92.351%. Figure 7.9 shows a decrease in the average number of ALCs from 40.666 to 37.266 over five iterations. The average fitness also decreased from 21.790 to 21.722. The average number of misclassifications increased from iteration one to three and then decreased at iteration five.

Table 7.15: *Building-window-float* - Negative selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
53	25	37.913 (±4.368)	27.353 (±3.741)	1.233 (±2.144)	14.967 (±5.129)	21.977 (±0.122)	21.690 (±0.065)
	50	23.467 (±5.185)	16.600 (±4.251)	1.033 (±1.810)	28.533 (±9.634)	24.698 (±0.210)	19.175 (±0.272)
	75	36.100 (±9.191)	27.433 (±7.420)	1.367 (±1.608)	26.567 (±6.956)	26.454 (±0.196)	16.549 (±0.256)
	100	56.100 (±13.547)	43.087 (±10.538)	1.133 (±1.871)	23.700 (±7.720)	26.953 (±0.169)	15.319 (±0.329)
106	**25**	**40.444 (±4.020)**	**25.233 (±2.644)**	**1.367 (±1.974)**	**14.167 (±4.662)**	**21.946 (±0.146)**	**21.735 (±0.051)**
	50	25.767 (±6.106)	16.289 (±4.260)	1.200 (±1.472)	27.000 (±7.883)	24.675 (±0.169)	19.229 (±0.180)
	75	35.933 (±8.158)	23.189 (±5.389)	1.400 (±1.773)	25.733 (±8.718)	26.424 (±0.184)	16.597 (±0.306)
	100	53.400 (±12.041)	34.856 (±8.013)	1.200 (±1.846)	23.233 (±6.095)	26.925 (±0.190)	15.334 (±0.261)
160	25	40.100 (±4.318)	19.500 (±2.125)	1.433 (±1.960)	15.467 (±4.939)	21.857 (±0.141)	21.762 (±0.052)
	50	25.667 (±5.996)	12.783 (±3.019)	1.167 (±2.019)	26.867 (±10.126)	24.702 (±0.237)	19.198 (±0.182)
	75	35.300 (±9.855)	17.633 (±4.936)	1.067 (±1.660)	28.367 (±7.522)	26.436 (±0.227)	16.521 (±0.326)
	100	56.267 (±19.490)	28.117 (±9.749)	1.133 (±1.634)	23.967 (±8.736)	26.986 (±0.226)	15.280 (±0.359)
214	25	39.667 (±2.832)	0.000 (±0.000)	1.367 (±1.991)	16.800 (±4.759)	21.841 (±0.185)	21.778 (±0.071)
	50	24.467 (±7.171)	0.000 (±0.000)	1.067 (±1.639)	28.800 (±8.290)	24.702 (±0.247)	19.195 (±0.218)
	75	30.700 (±6.914)	0.000 (±0.000)	1.167 (±1.840)	28.967 (±8.977)	26.444 (±0.219)	16.493 (±0.282)
	100	54.533 (±14.498)	0.000 (±0.000)	1.367 (±1.629)	24.800 (±6.150)	26.945 (±0.246)	15.332 (±0.380)

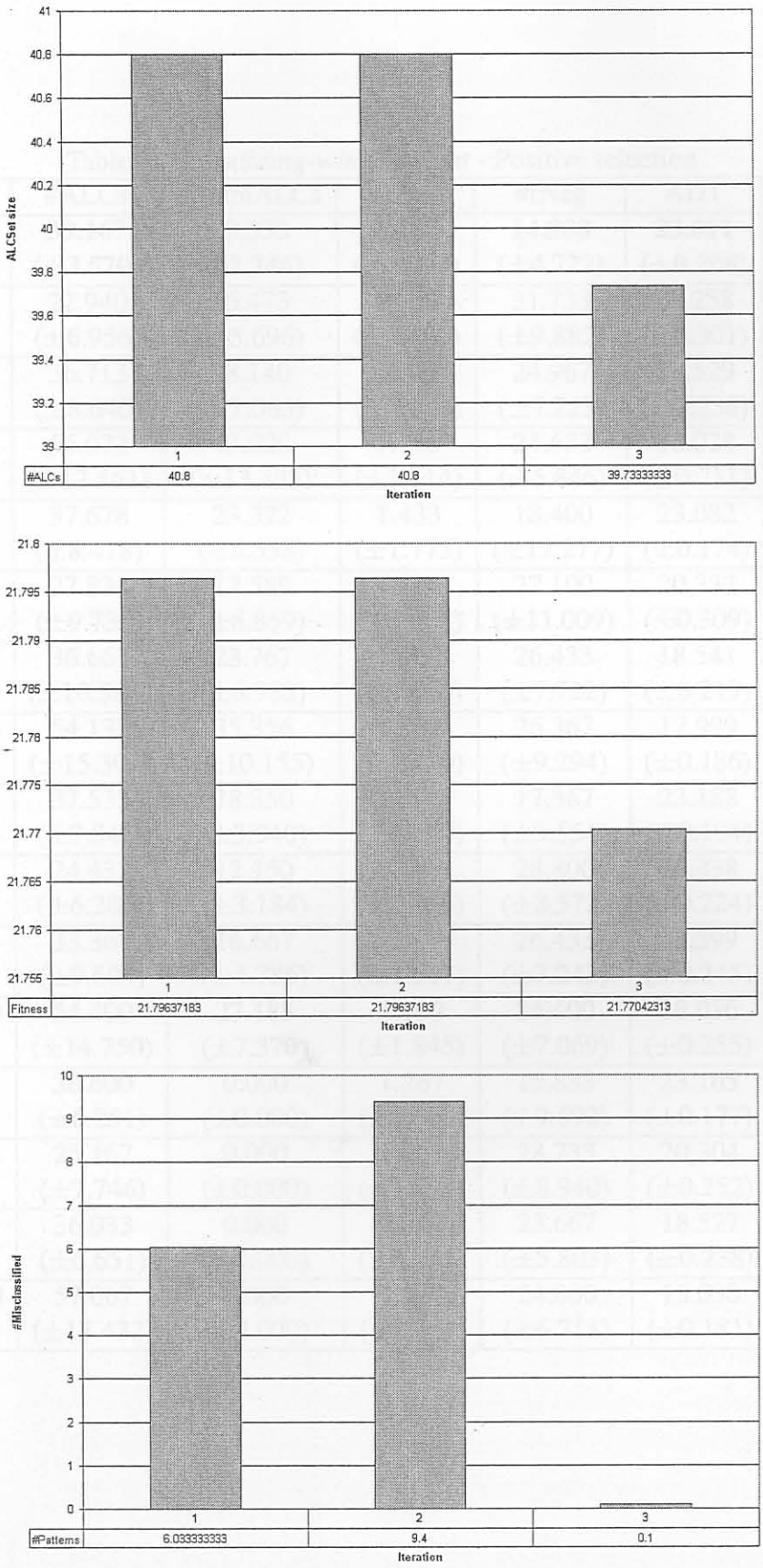


Figure 7.8: *Building-window-float* - Negative selection with IS=106 and W1=25

Table 7.16: *Building-window-float* - Positive selection

IS	W1	#ALCs	#MemALCs	#fPos	#fNeg	ADT	HD
53	**25**	**39.167 (±3.570)**	**28.333 (±2.746)**	**1.467 (±1.978)**	**14.900 (±4.722)**	**23.011 (±0.208)**	**21.681 (±0.099)**
	50	22.940 (±6.956)	16.473 (±5.696)	1.133 (±1.479)	31.733 (±9.882)	20.258 (±0.301)	19.088 (±0.308)
	75	36.713 (±8.640)	28.140 (±7.063)	1.033 (±2.008)	24.967 (±7.223)	18.529 (±0.236)	16.533 (±0.363)
	100	56.073 (±17.161)	43.020 (±13.433)	1.133 (±1.814)	24.633 (±6.846)	18.028 (±0.211)	15.330 (±0.219)
106	25	37.678 (±8.418)	23.322 (±5.558)	1.433 (±1.775)	18.400 (±12.277)	23.082 (±0.174)	21.691 (±0.187)
	50	27.833 (±9.739)	17.589 (±6.859)	1.100 (±1.863)	27.100 (±11.009)	20.332 (±0.309)	19.225 (±0.217)
	75	36.667 (±10.584)	23.767 (±6.988)	1.100 (±1.668)	26.433 (±7.722)	18.541 (±0.219)	16.511 (±0.270)
	100	54.133 (±15.301)	35.356 (±10.155)	1.367 (±1.790)	26.367 (±9.294)	17.999 (±0.186)	15.271 (±0.290)
160	25	37.533 (±7.847)	18.350 (±3.940)	1.300 (±1.664)	17.367 (±9.554)	23.188 (±0.194)	21.772 (±0.063)
	50	24.433 (±6.285)	12.150 (±3.184)	1.067 (±1.856)	28.800 (±8.572)	20.338 (±0.224)	19.214 (±0.190)
	75	33.367 (±9.604)	16.667 (±4.786)	0.900 (±1.517)	26.433 (±7.243)	18.599 (±0.245)	16.624 (±0.369)
	100	54.400 (±14.750)	27.183 (±7.370)	1.100 (±1.845)	24.400 (±7.069)	18.036 (±0.235)	15.338 (±0.278)
214	25	38.600 (±6.251)	0.000 (±0.000)	1.267 (±1.660)	15.833 (±9.692)	23.165 (±0.177)	21.754 (±0.056)
	50	25.167 (±7.746)	0.000 (±0.000)	1.067 (±1.799)	28.733 (±8.940)	20.304 (±0.252)	19.240 (±0.213)
	75	36.033 (±6.651)	0.000 (±0.000)	1.367 (±2.141)	23.667 (±5.803)	18.527 (±0.238)	16.512 (±0.246)
	100	57.067 (±15.432)	0.000 (±0.000)	1.167 (±2.001)	24.800 (±6.718)	18.038 (±0.181)	15.369 (±0.269)

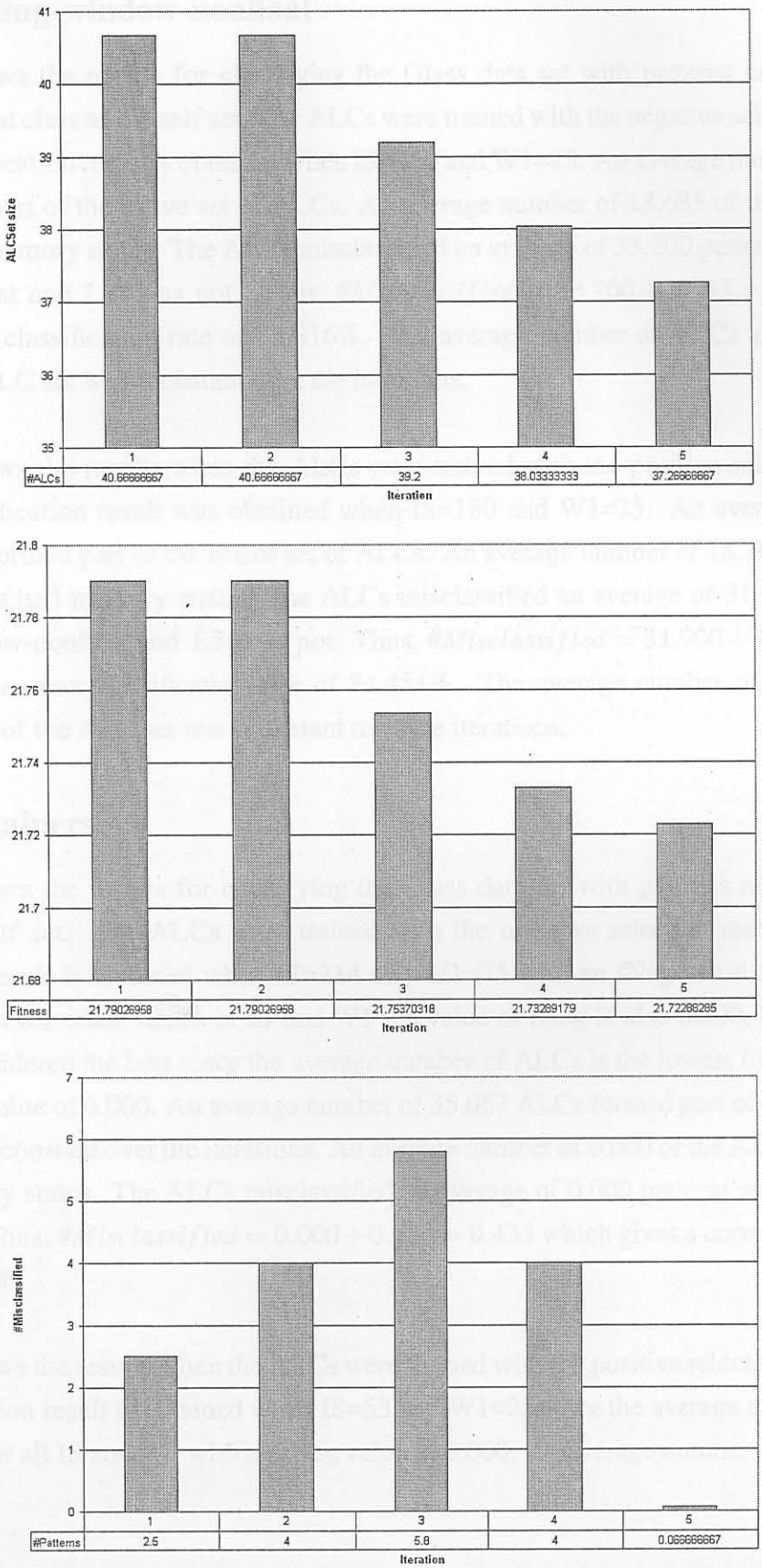


Figure 7.9: *Building-window-float* - Positive selection with IS=53 and W1=25