GA Optimisation of a Diagnostic Technique for Dementing Diseases based on Regional CSF Volume Measurements

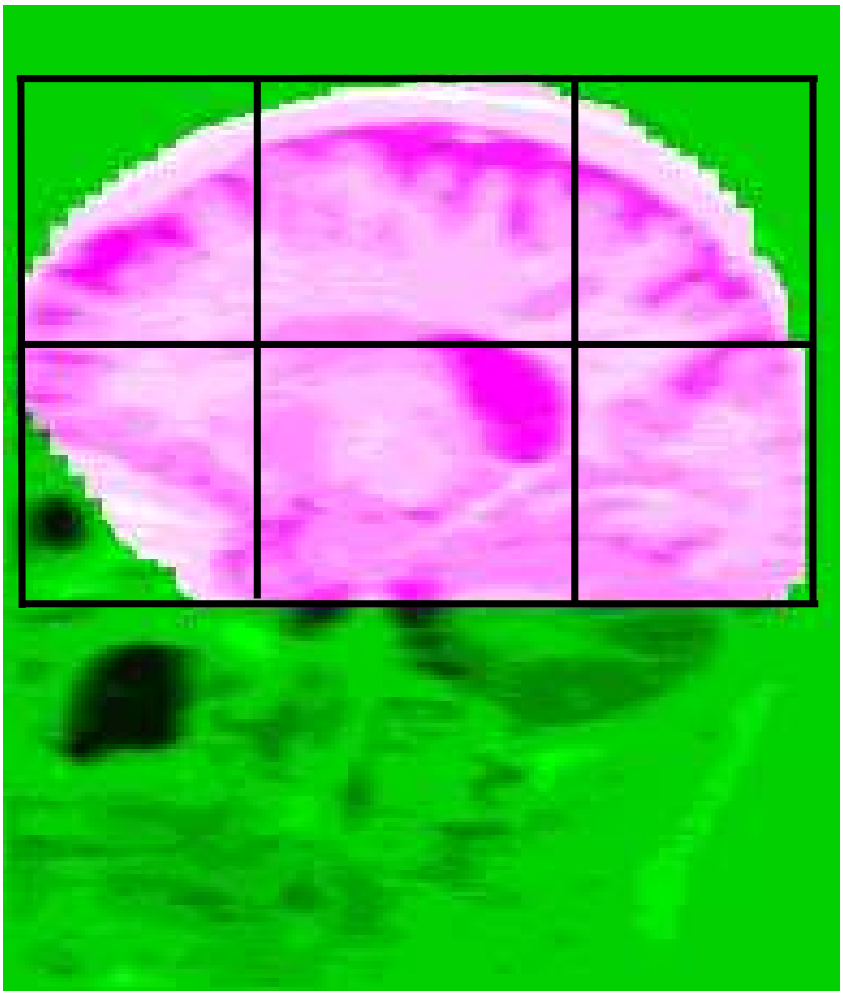
Dr. Paul A. Bromiley
Imaging Science and Biomedical Engineering

Overview

- Diagnosis of dementing diseases using the distribution of cerebral atrophy
  - use T2 IR MRI sequences
  - measure regional CSF volumes
- Optimisation of the technique using a genetic algorithm
  - GA primer
  - diversity preservation
  - multiobjective optimisation
  - preliminary results
- Further work
  - diagnosis
  - modelling

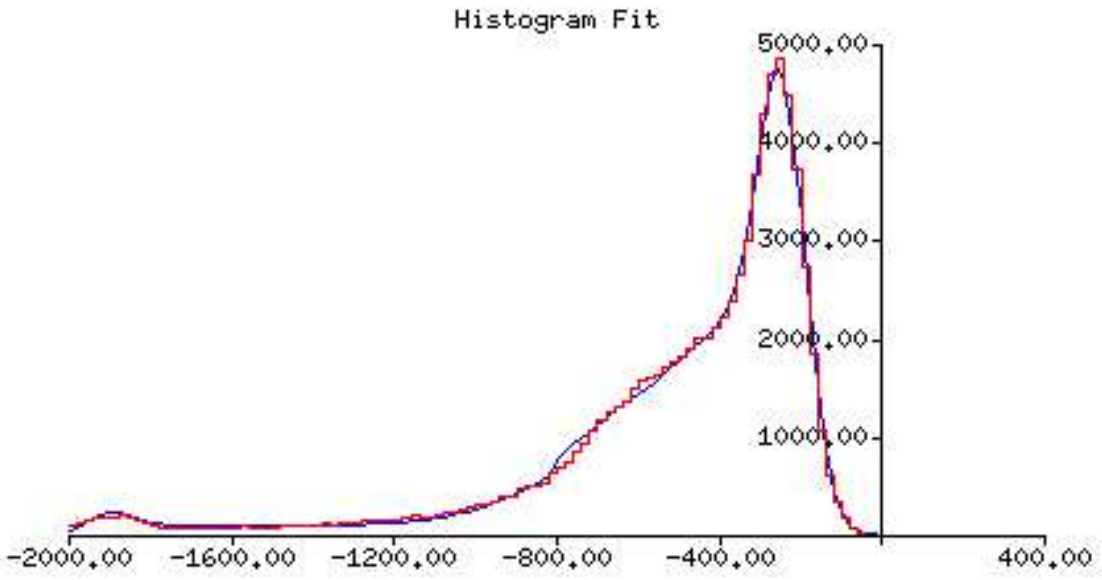
Atrophy Analysis: Introduction

- Original technique developed by Thacker et. al. (2002) Radiology 224 p278-285
  - hypothesis: dementing diseases result in specific patterns of cerebral atrophy
  - use minimal parameter atrophy measurement
  - co-register to a standard coordinate system
  - divide head into twelve equi-sized boxes



Atrophy Analysis: CSF Segmentation

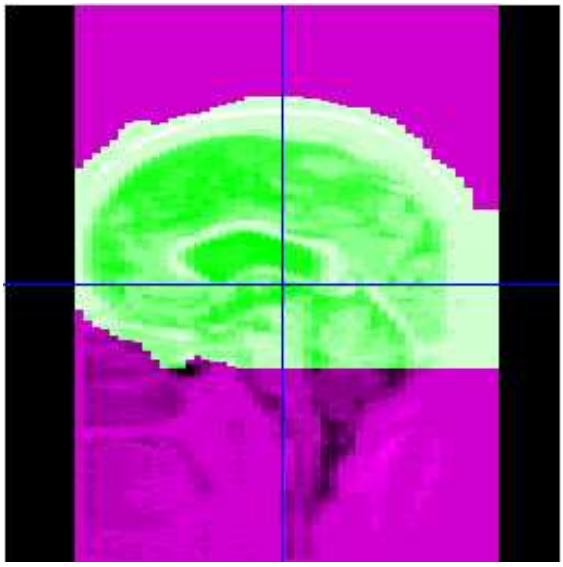
- Segment the CSF
  - use T2 IR scans
  - plot histogram of interior of head
  - fit with Gaussians + triangular distributions convolved with Gaussians

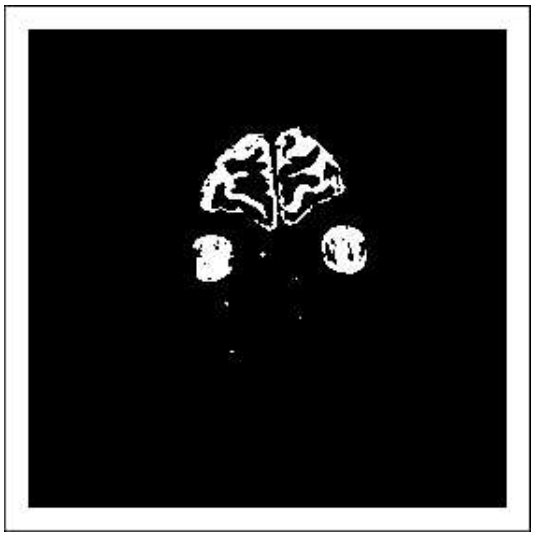


  - Vorkurka et. al., Am. J. Neuroradiology, 23 p. 459-467, 2002
  - place threshold at CSF probability = 0.5
  - produce binary CSF maps

Atrophy Analysis: Masking

- Binary images include all fluid spaces
  – CSF, eyes, sinuses
- Delete eyes, sinuses etc. using binary masks

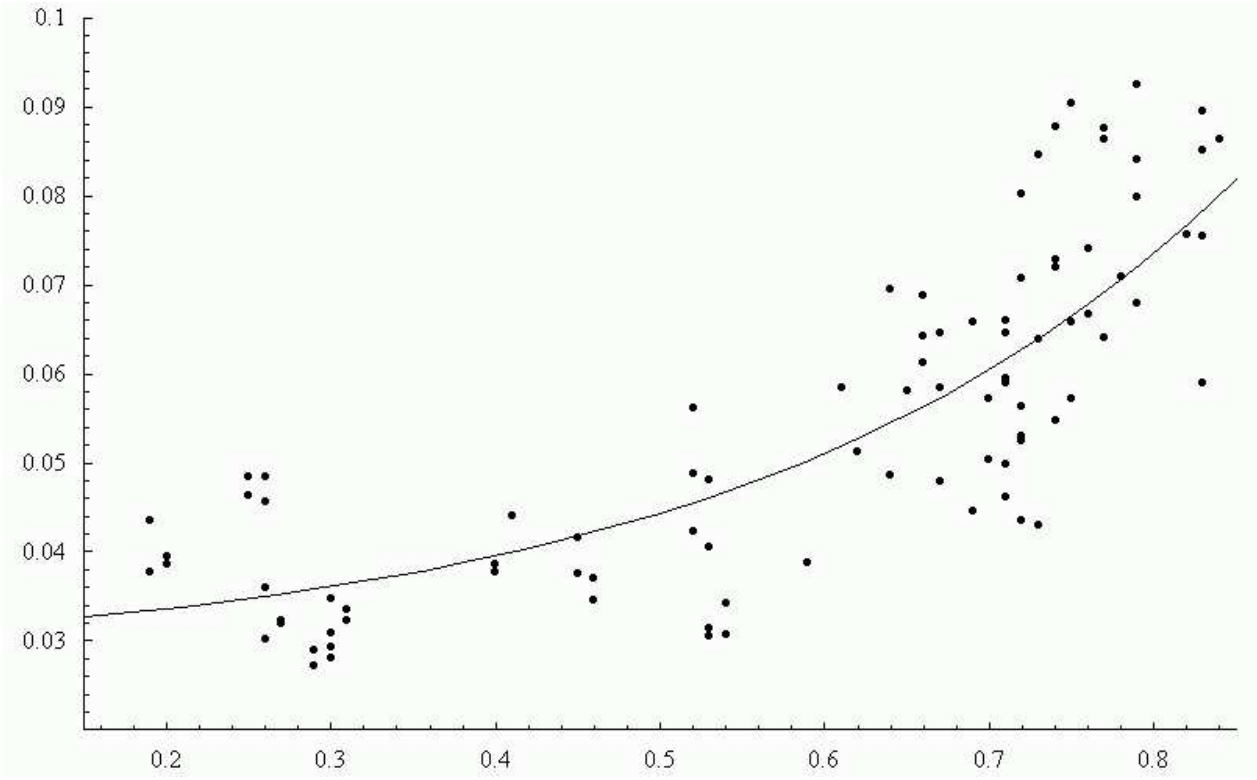




- Masks also define lower extent of the space: all other limits defined by outside of head

Atrophy Analysis: Age Correction

- Count CSF in the twelve boxes
- Normalise for head size (divide by total volume)
- Normalise for age-related atrophy



- Age normalisation
  - 94 normals: age range 19 to 85
  - fit curves

$$volume = a + b\ exp^{\ c\ age}$$

  - normalise to mean age point

Atrophy Analysis: Variable Reduction

- Calculate reduced variables
  - F = sum front four volumes
  - M = sum middle four volumes
  - B = sum back four volumes
  - P = sum left six volumes
  - S = sum right six volumes
  - U = sum top six volumes
  - L = sum bottom six volumes
- Calculate relative variables

$$W_1 = \frac{\sqrt{M} - \sqrt{F}}{\sqrt{2}}$$

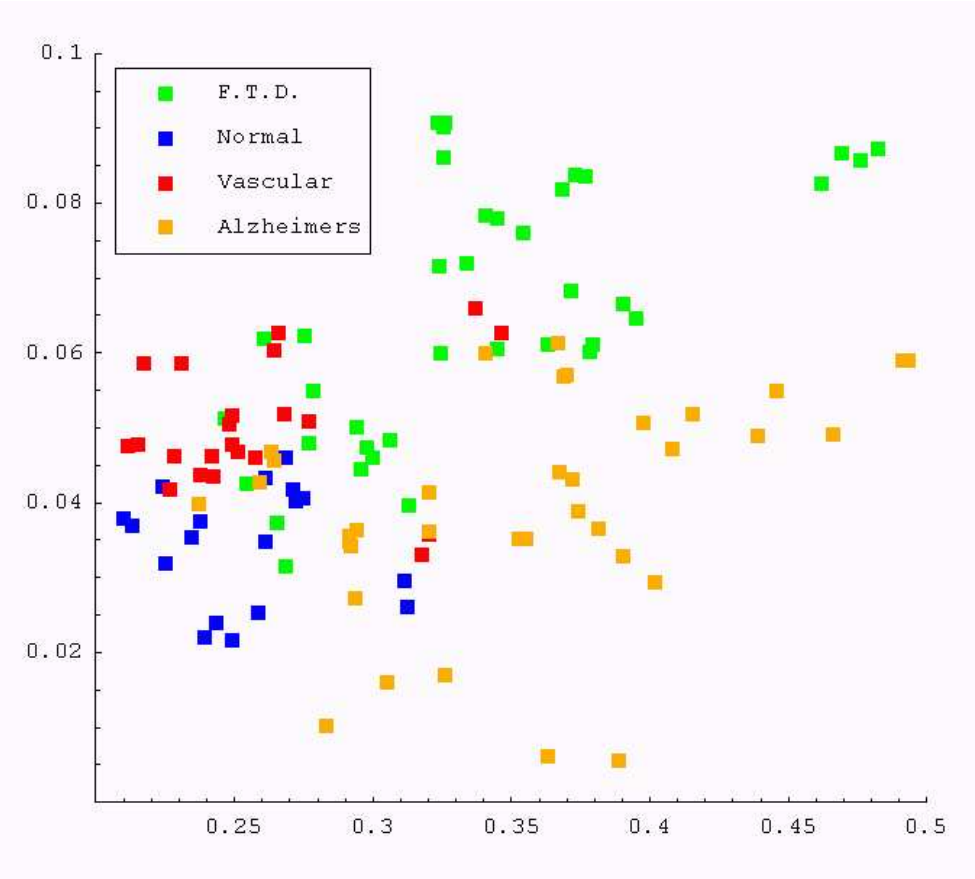
$$W_2 = \frac{\sqrt{M} - \sqrt{B}}{\sqrt{2}}$$

$$W_3 = \frac{\sqrt{F} + \sqrt{M} + \sqrt{B}}{\sqrt{3}}$$

$$W_4 = \frac{\sqrt{P} - \sqrt{S}}{\sqrt{2}}$$

$$W_5 = \frac{\sqrt{U} - \sqrt{L}}{\sqrt{2}}$$

Atrophy Analysis: Results



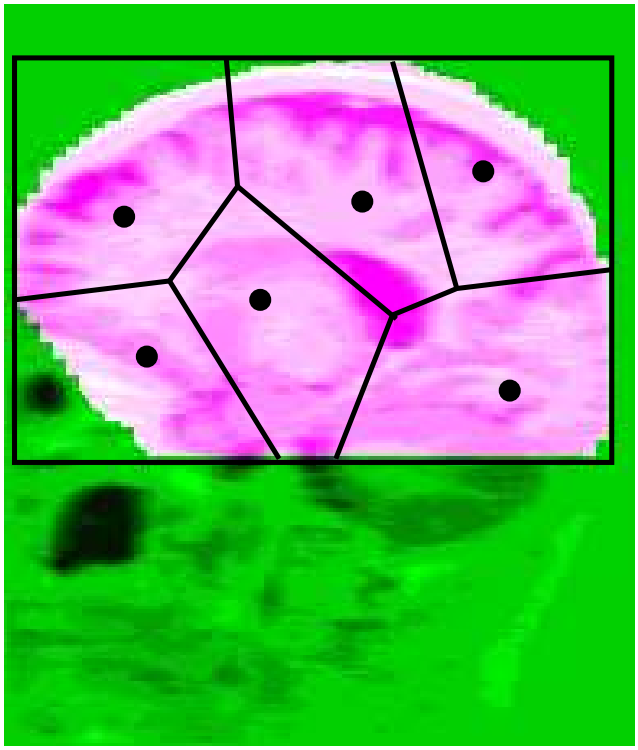
W_3 vs. W_2

- Cross-validated Parzen classifier results

Diagnosis	Normal	FTD	Vascular	Alz.
Normal	7	2	8	1
FTD	5	21	3	7
Vascular	3	2	13	4
Alz.	1	3	6	28

Motivation for GA Optimisation

- CSF boxes not related to brain structure
  - sub-optimal e.g. FTD often produces degeneration of frontal lobes
- Optimise boundary positions
  - replace boundaries with centre points
  - tessellate space with Voronoi cells



  - point pairs form natural building blocks
  - cost function noisy, form unknown
  - use GA

GA Primer

- Three-operator GA
  - code the parameters of the optimisation problem as a binary string
  - generate a population of random strings
- Reproduction
  - copy the strings according to fitness
- Crossover
  - pair the strings and swap genes between them
- Mutation
  - randomly flip some bits
- The Fundamental Theorem
  - "the frequency of highly-fit sub-strings grows exponentially, in proportion to the ratio of their average fitness to the population average"
- D.E. Goldberg. Genetic Algorithms: in Search, Learning and Machine Optimisation

GA Example

- Example: optimise x^2 over x=0 to 31
- Code parameters as five bit unsigned integer
  - x=0 at 00000, x=31 at 11111
- Population=4

String	x	f(x)	$\frac{f_i}{f}$	Expected count	Actual count
01101	13	169	0.14	0.58	1
11000	24	576	0.49	1.97	2
01000	8	64	0.06	0.22	0
10011	19	361	0.31	1.23	1

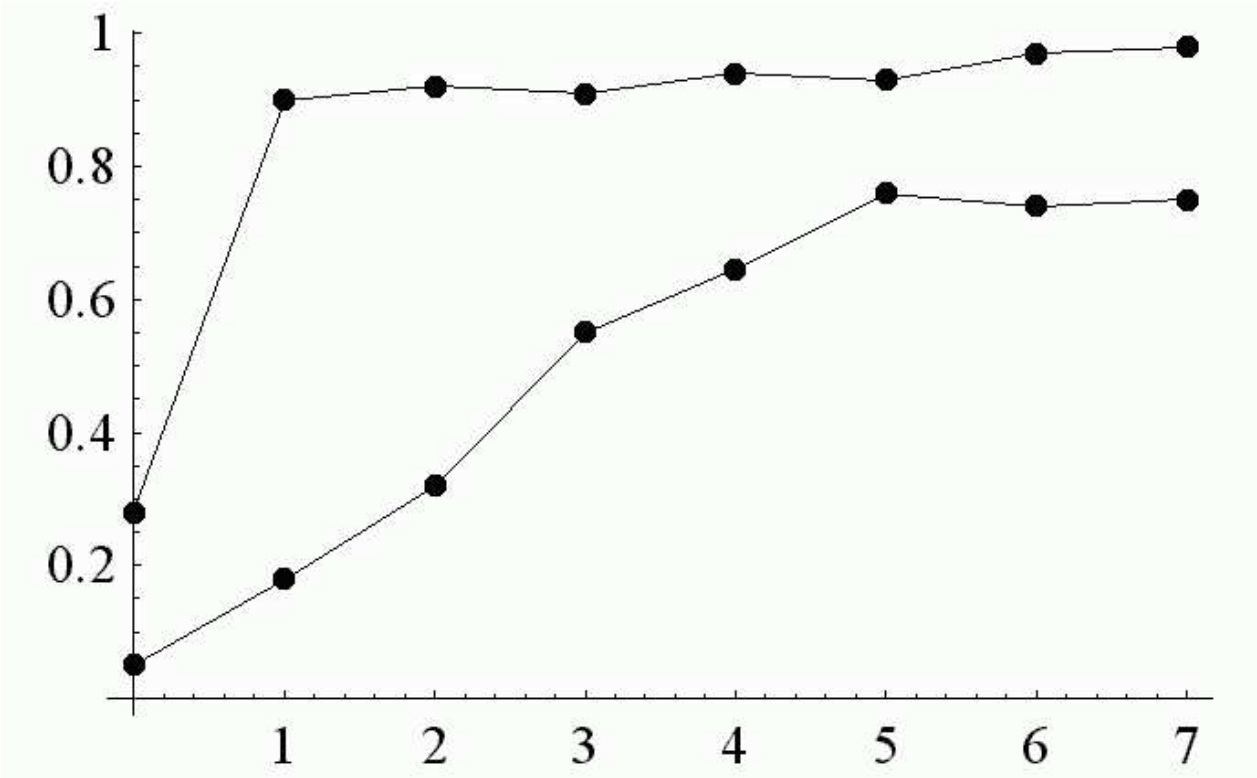
- Average fitness = 293 : Max fitness = 576

MP	Mate	Crossover site	New pop	x	f(x)
0110—1	2	4	01100	12	144
1100—0	1	4	11001	25	625
11—000	4	2	11011	27	729
10—011	3	2	10000	16	256

- Average fitness = 439: Max fitness = 729

Diversity Loss / Premature Convergence

- Simple example: $f(x) = (x/c)^{10}$
  - range x = 0 to 1: 30-bit binary encoding
  - $p_m = 0.03$, $p_c = 0.6$, $n = 30$ (De Jong, 1975)



Fitness vs. Generation

- Approaches optimum, but does not reach it
  - population is degenerate by generation 7
  - diversity loss
  - premature convergence

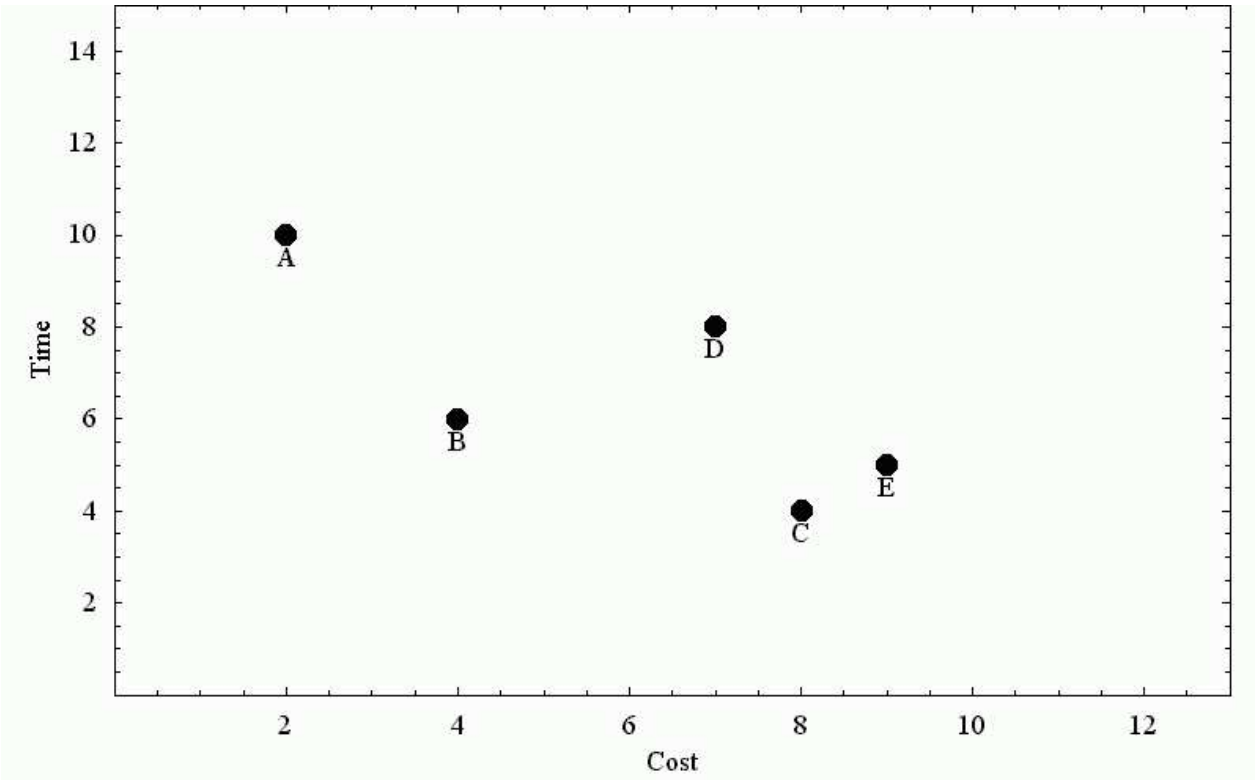
Diversity Preservation

- Most GA research focuses on diversity preservation
- Many schemes:
  - alternative selection schemes
  - fitness scaling
  - crowding and preselection
  - niching and speciation
  - mating restriction

but...

- No free lunch theorem:
  - all algorithms perform equally well over the set of all problems
- There is no such thing as the best GA
- Code can be freely downloaded from the web
  - IlliGAL: www-illigal.ge.uiuc.edu

Multi-objective Optimisation



- D, E: *dominated* solutions
- A, B, C: *non-dominated* solutions
- Non-dominated solutions form the *Pareto Front*
- GAs ideal:
  - multiple individuals in population
  - entire Pareto Front in a single run
  - diversity must be preserved

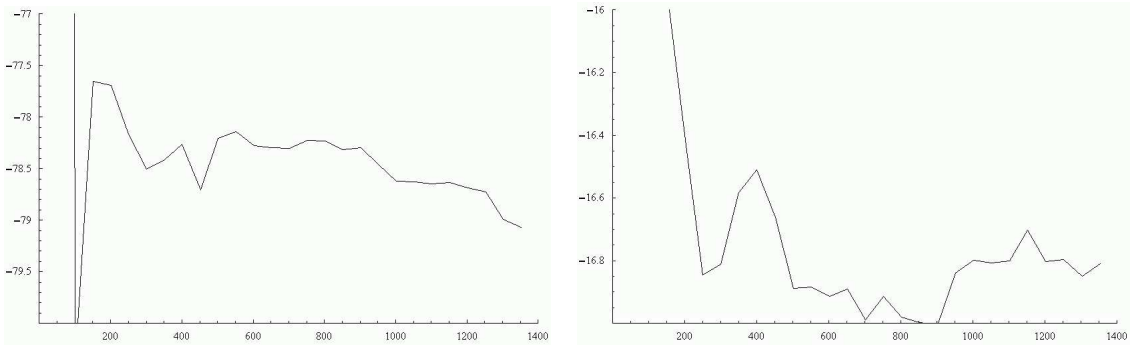
GA Goals

- Multi-objective cost function
  - use diagonal elements of confusion matrix
  - diversity preservation essential
- Diversity preservation
  - global convergence will occur if any pair of solutions can interact
  - nature: strict mating restrictions (inter-species) + dynamic cost function (intra-species)
  - convergence occurs locally, but not globally
- Minimise free parameters
  - population size, mutation probabilities
  - diversity preservation strategy parameters (e.g. crowding factor)
- NIAC GA: www.tina-vision.net/dosc/memos.php

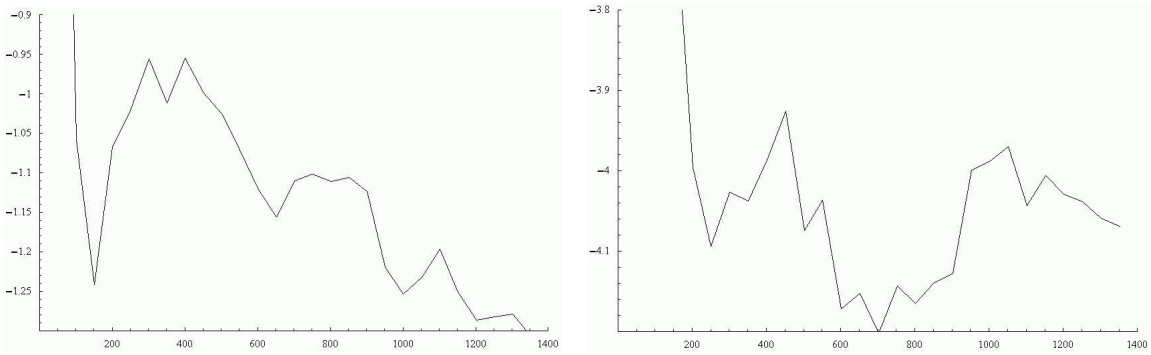
GA Optimisation: Preliminary Results

- Cost function: sum of cross-validated Parzen classifier probability over each disease group
- GA optimisation run for 1400 iterations ($\approx$ 900 hours)
  - 94 normals, 24 Alzheimers, 9 Vascular
  - 13 FTD, 6 Lewy, 15 Schiz
- Very unbalanced data set
  - result cannot be compared to original paper

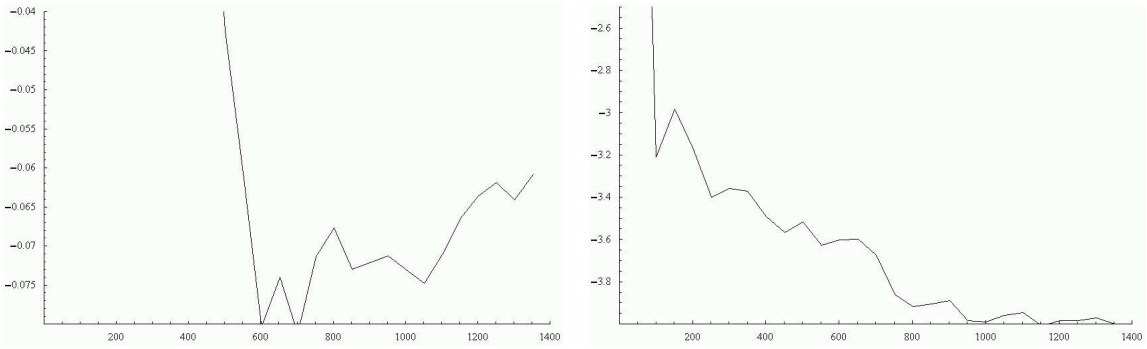
Extracted Set Cost vs. Time



Normals and Alzheimers Disease



Vascular Dementia and FTD



Lewy Body Disease and Schizophrenia

A Sample Solution 1

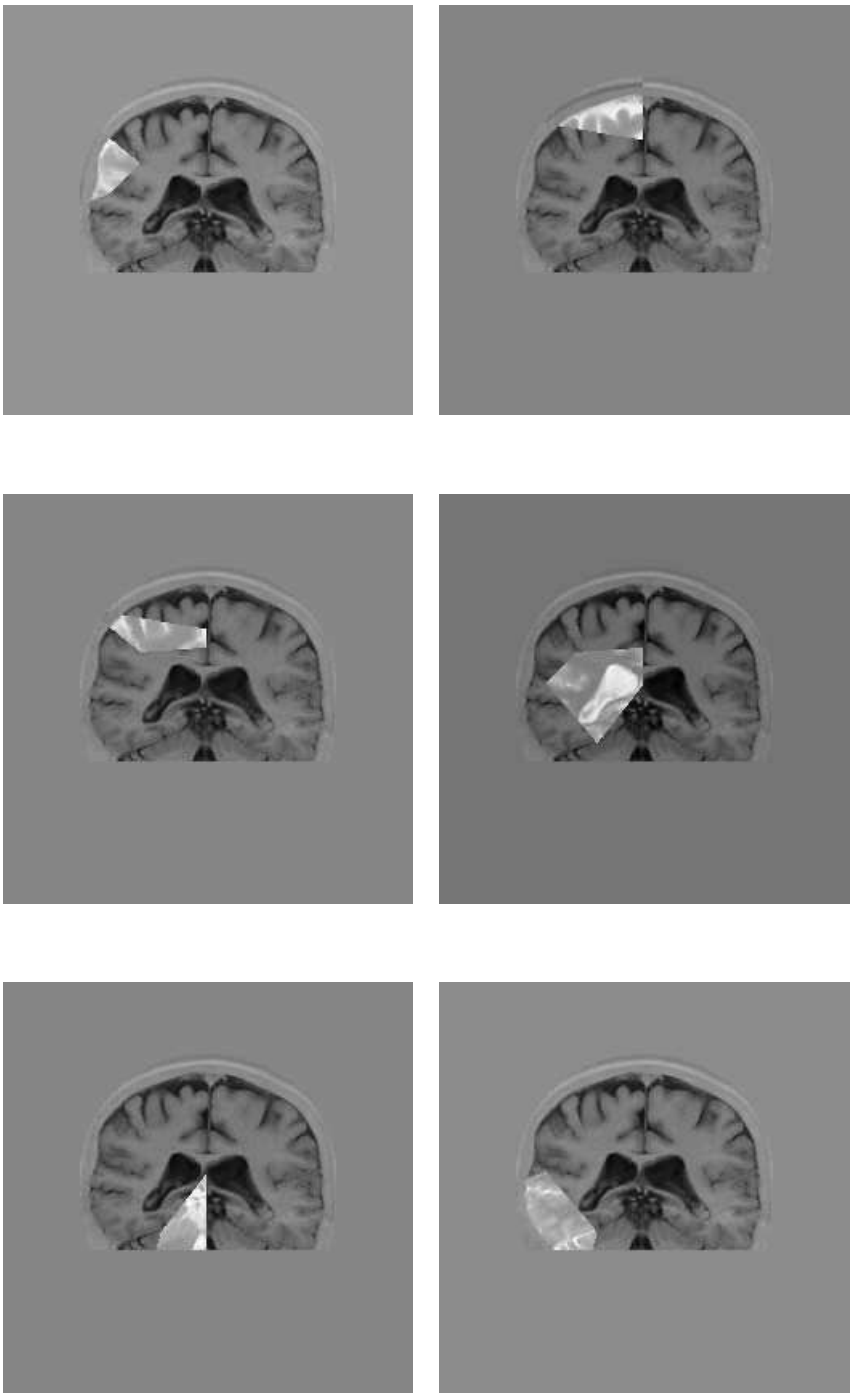
- Take solution with best averaged proportional cost from 1352 iteration extracted set

	Diagnosis					
	Norm.	Alz.	Vasc.	FTD	Lewy	Sciz.
Norm.	82	1	4	1	3	6
Alz.	2	21	0	4	2	0
Vasc.	4	0	4	0	0	3
FTD	0	1	0	5	1	0
Lewy	1	1	0	1	0	0
Schiz.	5	0	1	2	0	6
Total	94	24	9	13	6	15
C.F.	82	17	2	6	0	3

(Last row shows diagonal of matrix for original atrophy analysis method)

A Sample Solution 2

- What has happened to the CSF boundaries?



Conclusion

- Coarse regional measurements of CSF volumes are sufficient to diagnose atrophic diseases.
- Overall, the GA optimisation has improved the diagnostic capability of the atrophy analysis technique.
  - diagnostic improvements of $\approx 10\%$ should be expected
  - a balanced data set is required for the final optimisation.
- New decision boundaries focus on regions at the top of the frontal lobe, and separate out the ventricles
- Indicates more points, and therefore more data, could be used in future work.

Further Work: Extension to Cerebral Blood Flow Modelling

- Dementing disease diagnosis
  - use binary maps of CSF
  - determine optimal CSF measurement regions for diagnosis
- MIAS IRC Autoregulation project
  - modelling of cerebral blood flow
  - produce vector maps of blood flow
  - regional mean log flow correlates with disease
- $\Rightarrow$ Combination of these techniques
  - take diseases with a vascular component
  - use GA to determine optimal measurement regions
  - construct diagnosis systems for vascular diseases
- Optimise over any groups that may correlate with blood flow patterns
  - e.g. carotid stenosis vs. normal
  - non-disease groups

Further Work: Application to Modelling

- The optimised regions are those regions which show the biggest correlation with group membership
  - sensible division of the brain into compartments for blood flow modelling
- This will not:
  - produce maps of arterial territories
  - produce a "one-size-fits-all" solution
- This will:
  - find the most informative model compartments for the set of patient groups entered into the analysis

www.tina-vision.net