

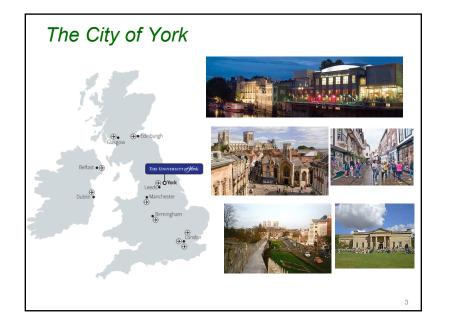
Welcome

- Stephen L. Smith received a BSc in Computer Science and then an MSc and PhD in Electronic Engineering from the University of Kent, UK. He is currently a Reader in the Department of Electronics at the University of York, UK.
- Steve's main research interests are in developing novel representations of evolutionary algorithms particularly with application to problems in medicine. His work is currently centered on the diagnosis of neurological dysfunction and analysis of mammograms.
- Steve is co-founder and organizer of the MedGEC Workshop, which is now in its twelfth year. He is also coeditor of a book on the subject (John Wiley, November 2010).
- Steve is associate editor for the journal Genetic Programming and Evolvable Machines and a member of the editorial board for the International Journal of Computers in Healthcare and Neural Computing and Applications.
- In 2013 Steve co-founded ClearSky Medical Diagnostics, a university spinout company for developing medical devices that exploit genetic programing technologies.



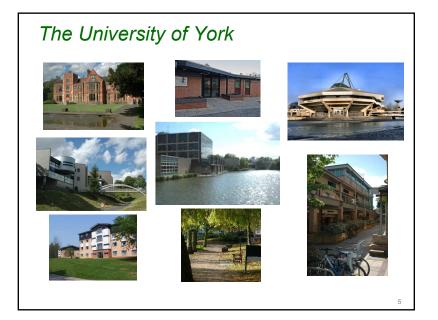


UNIVERSITY of York



The University of York





Overview

- Introduction to medical applications of EC
 - how these differ from other real-world applications
- Overview of previous work
 - from a medical and EC point of view
- Case examples of medical applications
- Practical advice on how to get started
 - choosing an application and obtaining good data
- Summary

Introduction

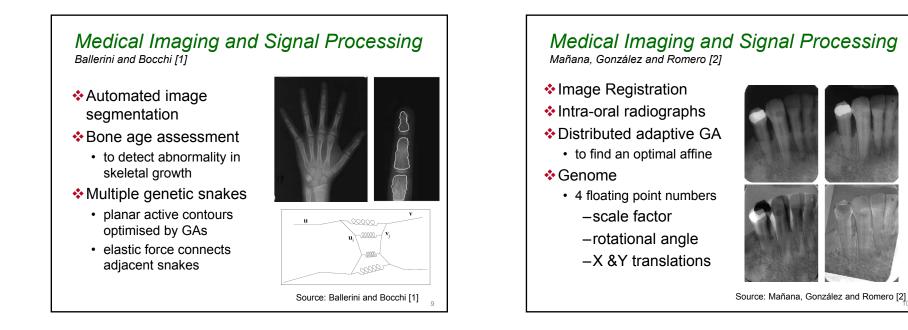
- Challenging application area
 - often no reliable "ground truth" or "gold standard"
 - · access to data difficult
 - multiple ethical considerations

Opportunities

- unexplored territory
- funding and publications
- make a difference

Previous Work

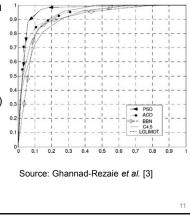
- Medical Imaging and Signal Processing
- Data Mining Medical Data and Patient Records
- Modelling and Simulation of Medical Processes
- Clinical Expert Systems and Knowledgebased Systems
- Clinical Diagnosis and Therapy

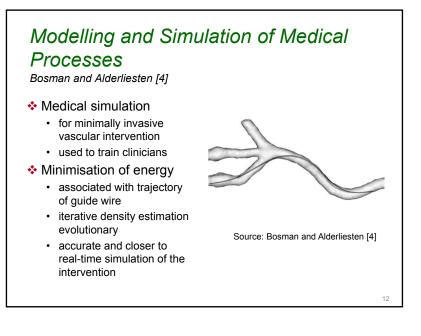


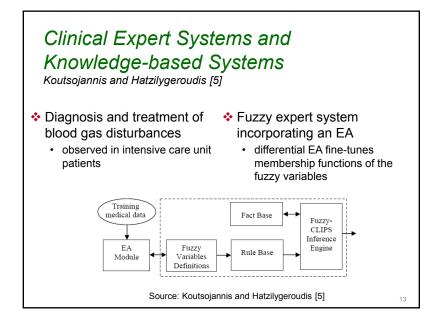
Data Mining Medical Data and Patient Records

Ghannad-Rezaie et al. [3]

- Surgery candidate selection
 - · for temporal lobe epilepsy
 - integrates a classifier with a particle swarm algorithm (PSO)
- Compared with:
 - ant colony optimisation (ACO)
 - Bayesian Belief Network (BBN)
 - C4.5 (a decision tree approach)
 - LOcal Linear MOdel Tree (LOLIMOT)

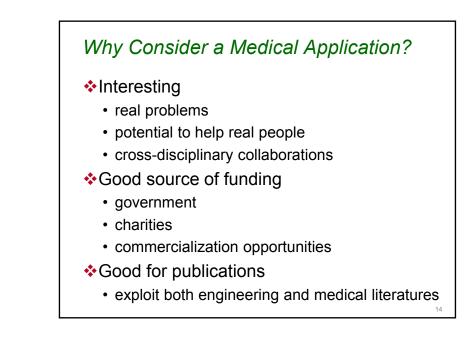






Choosing a Medical Application

- Review the field
 - through the medical and engineering literature
- Choose a novel application area
 - with a clinical need (medical or financial!)
- Contact a health professional
 - go and talk to them
- Evaluate the scope for obtaining data
 - and a "gold standard" or "ground truth"



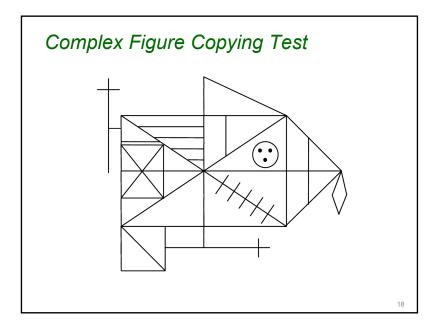
Datasets

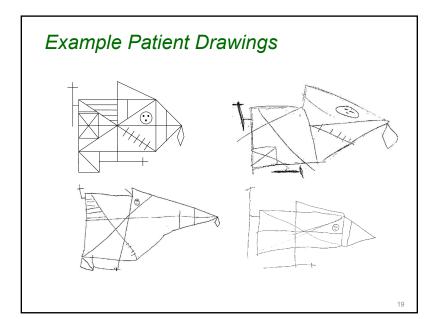
- The performance of any evolutionary algorithm is only as good as the data it is trained on
 - often no definitive clinical test for condition under investigation
 - datasets usually involve subjective clinical assessment
 - often have missing or corrupted values
 - resulting in insufficient good data for statistical
 - can be ameliorated by the use of k-fold crossvalidation

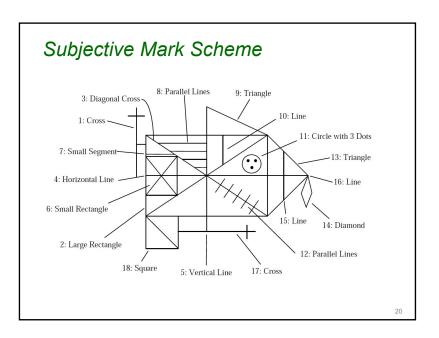
Datasets

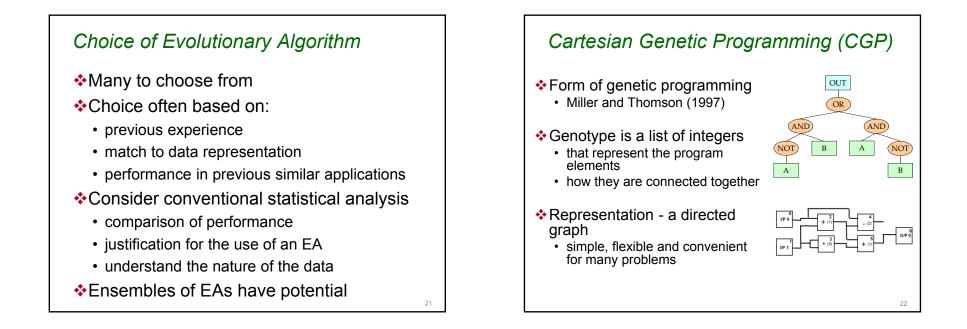
- When subjective evaluation is involved:
 - engage multiple assessors
 - to provide greater confidence
 - agree a consistent marking protocol

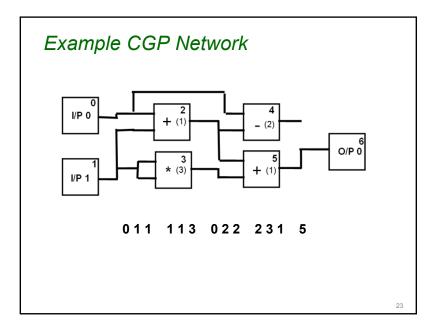
 several standards usually in use
 - · gather data from multiple sites
 - to compensate for local marking practice
 - $\, \text{to} \ \text{sample representative demographic population}$
- Engage a health statistician before you start

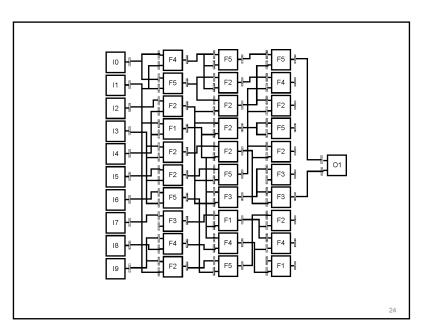


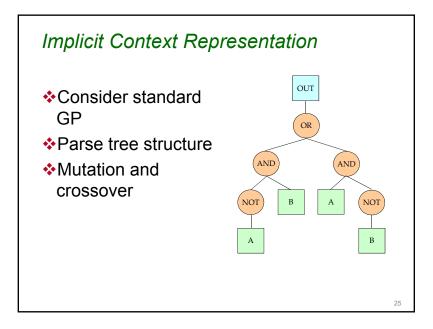


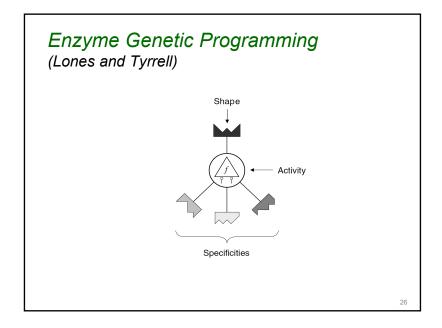


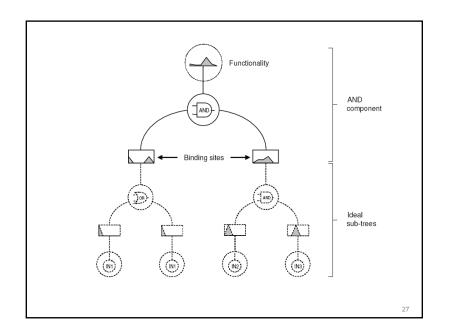


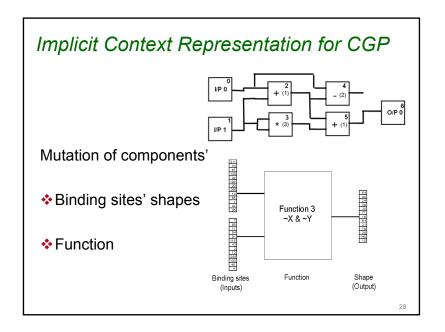


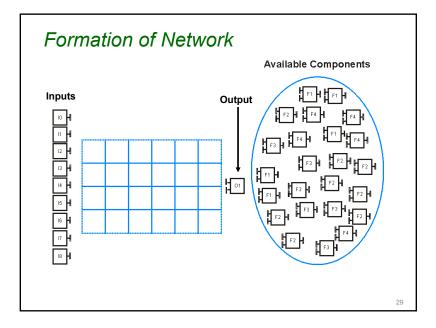


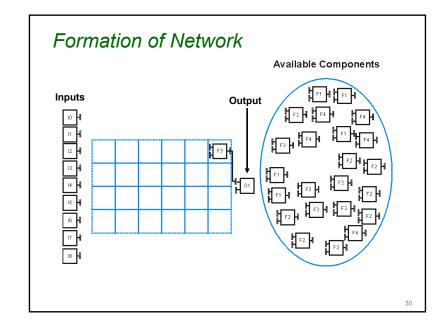


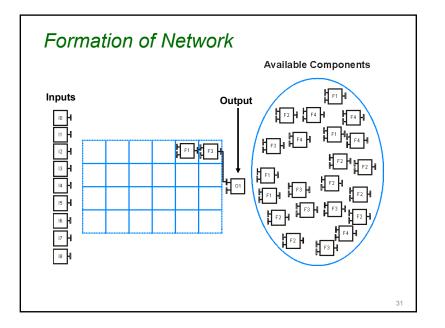


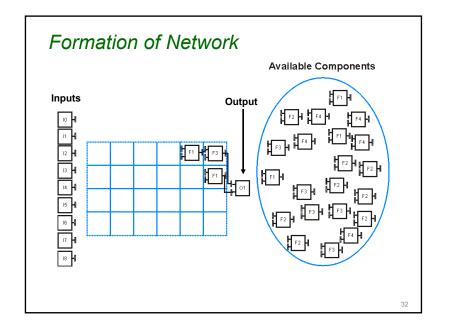


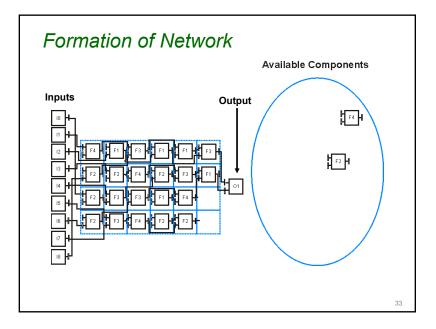


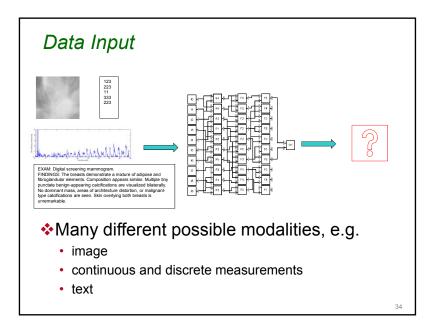


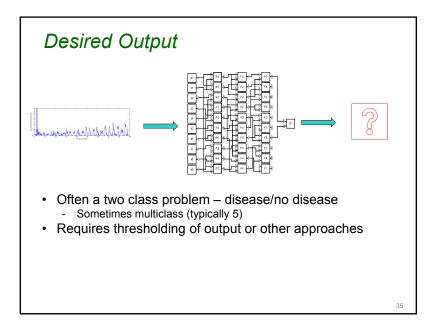


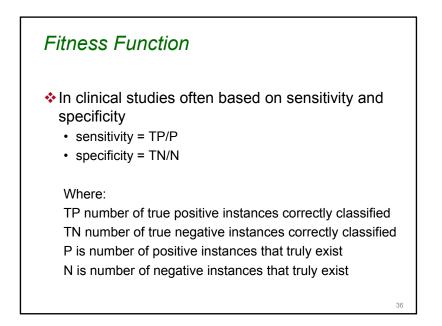


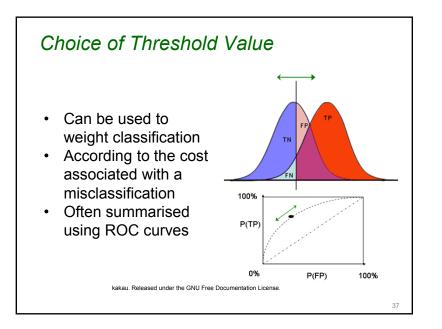


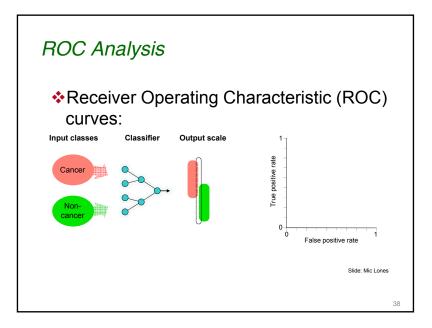


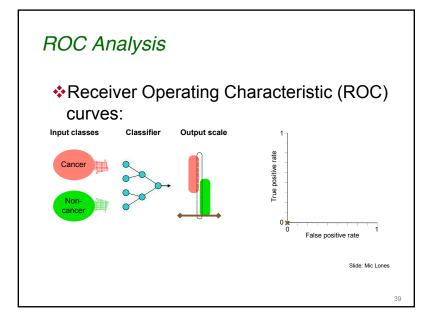


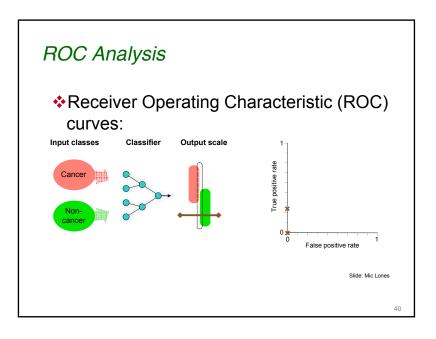


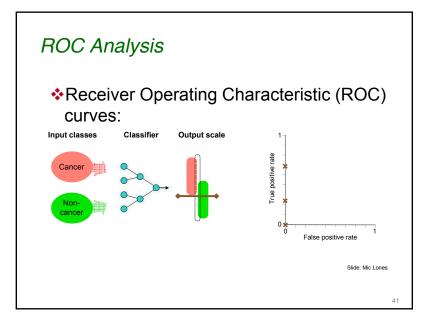


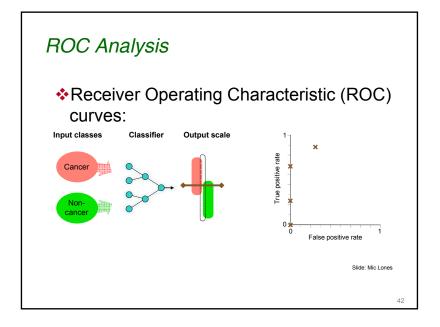


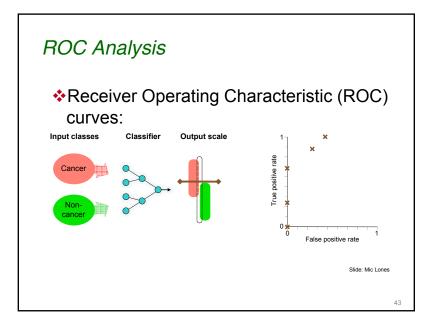


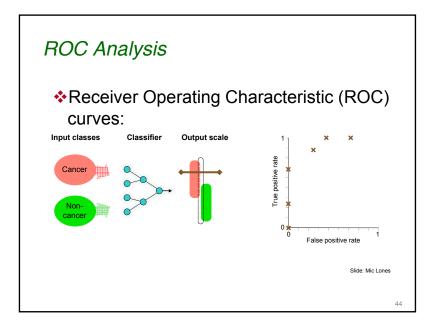


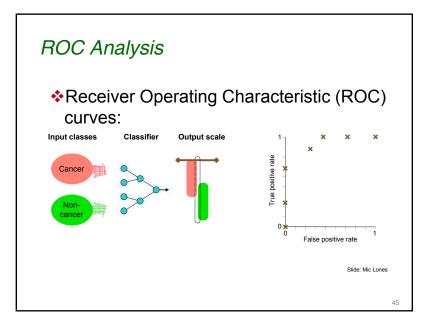


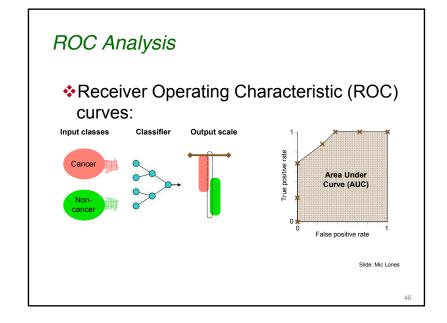


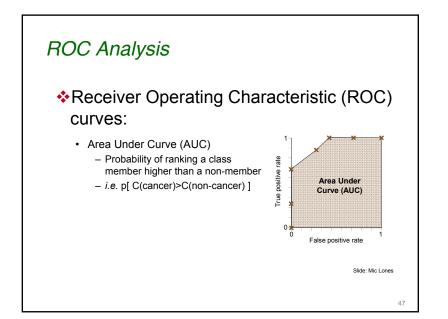


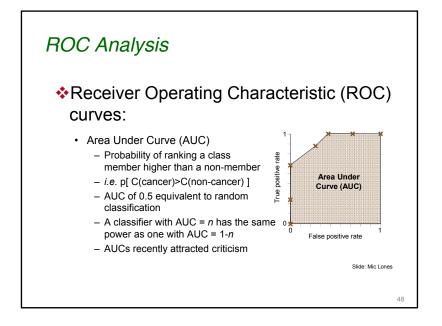


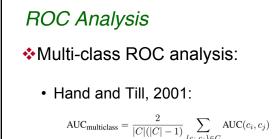






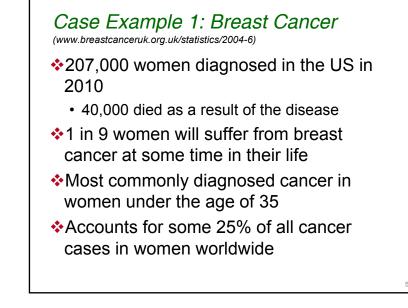


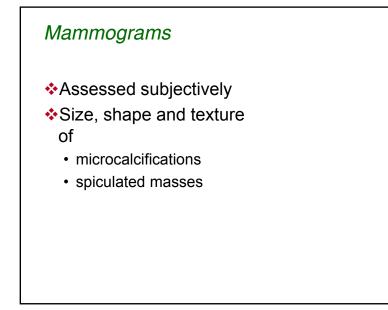




- ≈ mean of AUCs for each pair of classes
- Direction of AUCs (> or < 0.5) is important
 - -i.e. must separate classes in the correct order

D. Hand and R. Till. A simple generalization of the area under the ROC curve to multiple class classification problems. *Machine Learning*, 45(2):171-186, 2001.



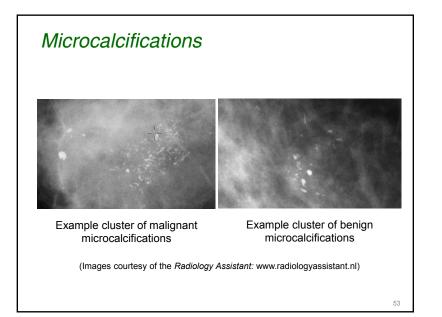


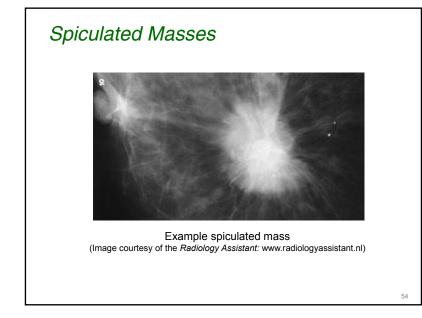
Microcalcifications

- Calcium deposits
 - · secretions from ductal structures

52

- Occur in clusters
- 40-50% cancerous
- Discriminated by:
 - shape
 - size
 - texture
 - distribution





Datasets

- Several publically available datasets of mammograms
 - Mammographic Image Analysis Society (MIAS) Database
 - University of South Florida Digital Mammography Database (USFDMD)
 - Lawrence Livermore/University of California (LLNL/UCSF) Database

Many privately constructed datasets

Datasets

- All have their disadvantages
- Lack consistency in:
 - classification
 - · demographic spread
 - diagnosis
 - supporting diagnostic information

56

- digitization
- resolution
- file format

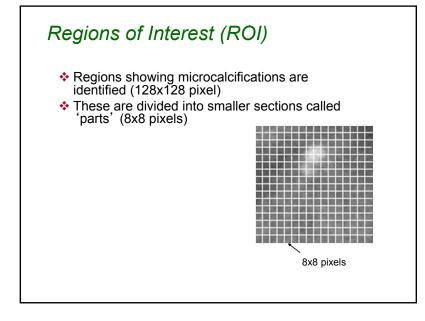
Using Existing Datasets

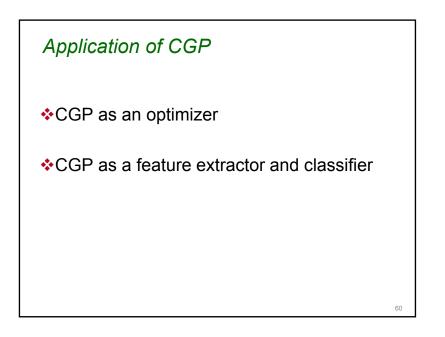
Using other peoples datasets may have implications!

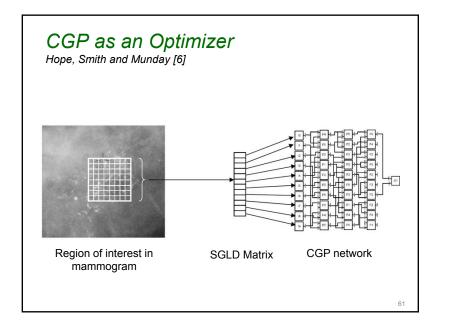
LLNL/UCSF database

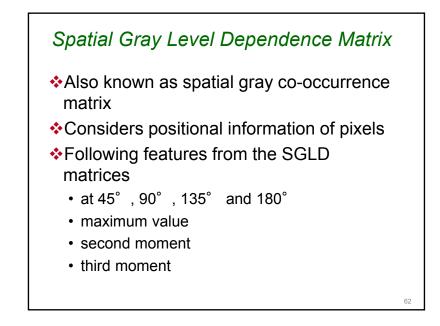
From "Somewhere in England"

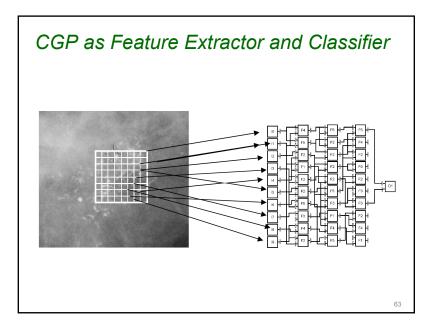
Constructing a New Dataset Requires: clinical cooperation authority suitable software tools especially, the user interface

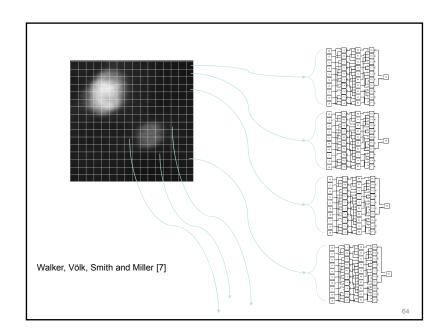


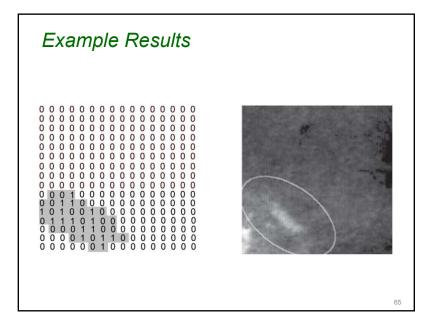


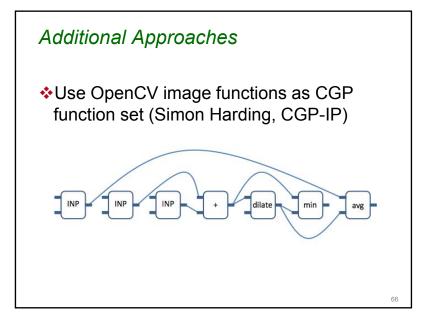












Case Example 2: Parkinson's Disease

- Common, chronic, progressive neurodegenerative brain disease
- Affects the control of muscles
- Incidence of 1 person in 1000
 - about 1 person in 100 over the age of 60

Insufficient formation and action of dopamine

Primary Symptoms of Parkinson's Disease

68

- Tremor
 - 3Hz-8Hz
- Rigidity
 - cogwheel
- Bradykinesia
 - slowing in movement
- Postural instability
 - poor balance

Conventional Diagnosis of Parkinson's Disease

- Diagnosis of idiopathic Parkinson's Disease
 - · based on clinical features
 - · often very subjective
 - poor sensitivity up to 25% error in diagnosis
 - PET scans to detect decreased dopamine activity

Medication

- L-dopa, artificial dopamine
- · has limited life due to feedback inhibition
- eventually becomes counterproductive.

Computer-Based Assessment

- Based on neuropsychological tasks
 - figure copying tasks
 - finger tapping tasks
 - measurement of tremor at rest
- Patient' s movements digitised in real-time

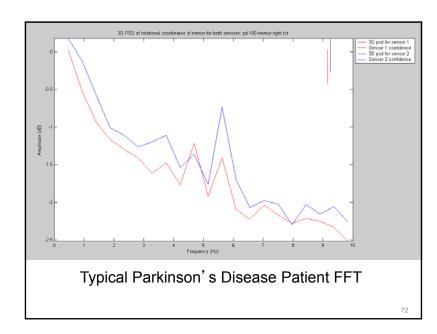


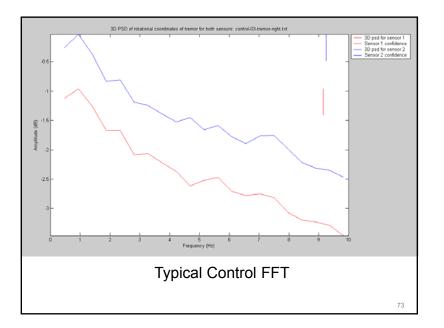


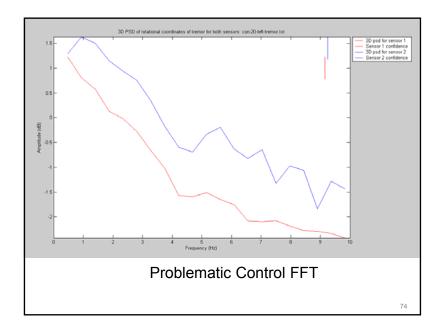
Data Analysis

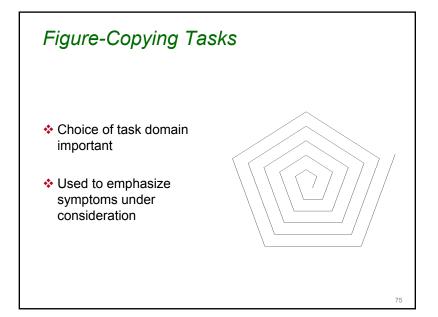
Initially, results poor

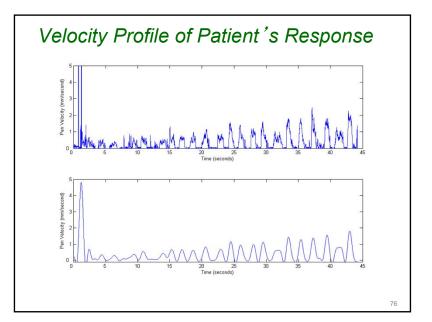
- Closer inspection of patient data
 - · patients with Parkinson's disease
 - controls with no known neurodegenerative condition



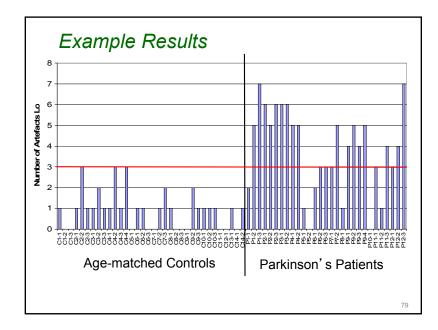


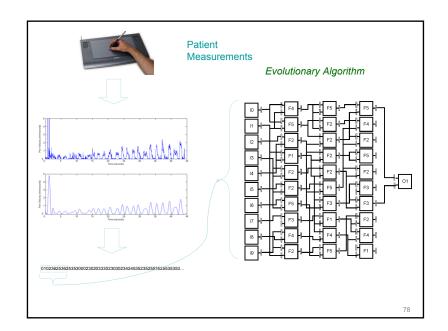




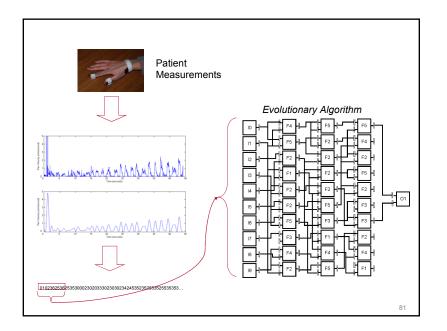


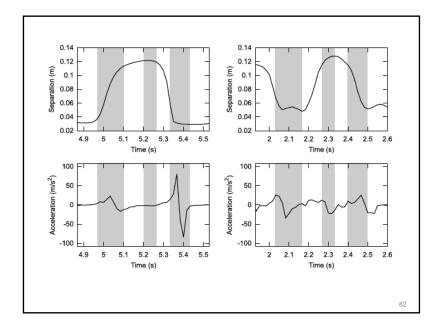
<text>

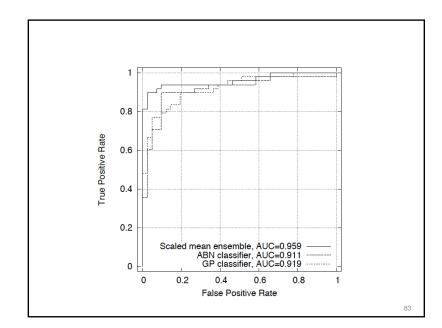


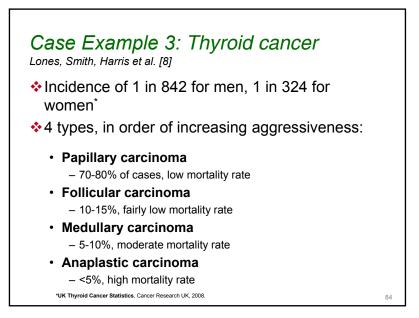


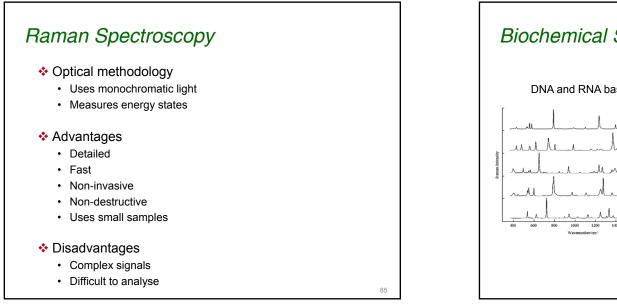


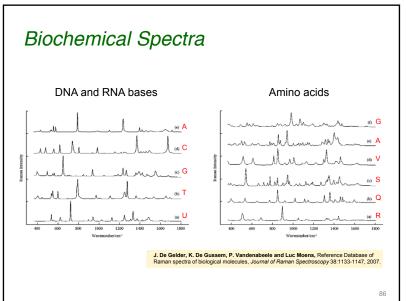


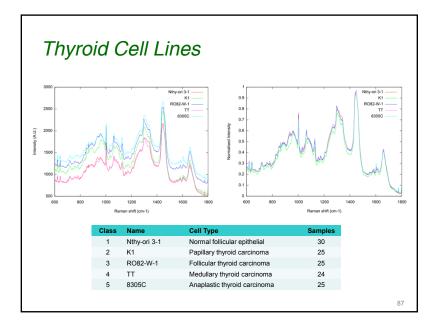


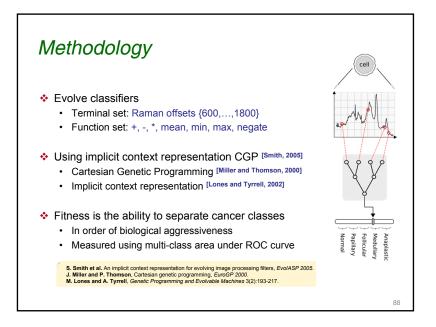


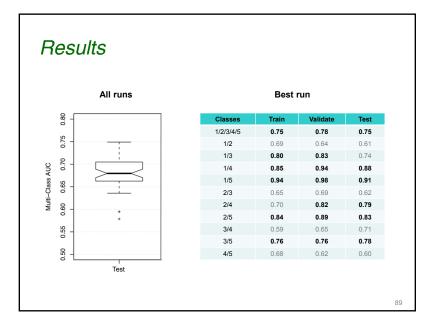


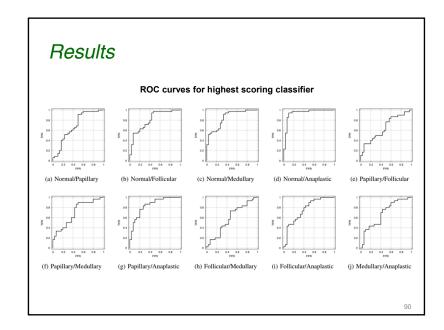








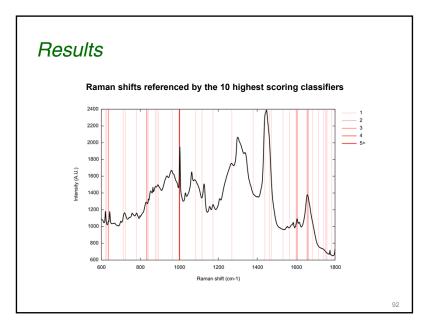


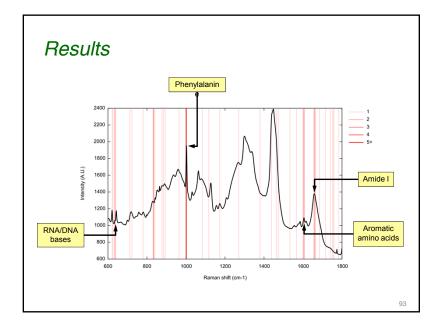


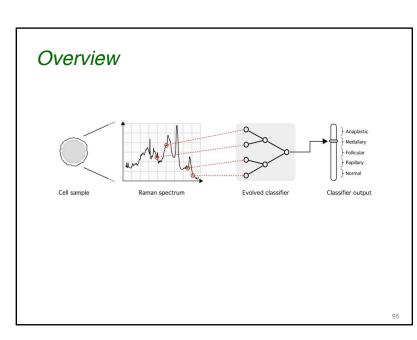
Results

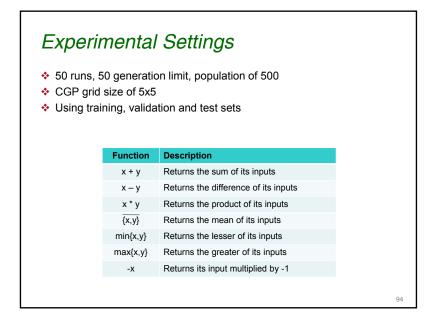
The 10 highest scoring classifiers

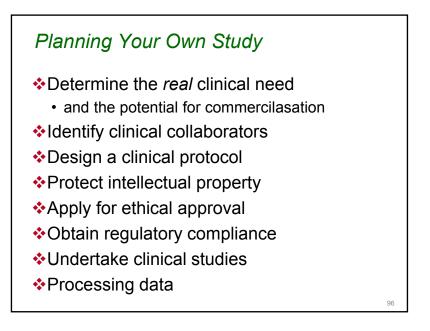
Expression	Test AUC
$out = \overline{\{885, 1001\}} - 1607 - \max\{1656, 1117\}$	0.75
$out = (841 + 1002) - (836 + 1600) + (max\{-622, 1657 * 1655\} - 1657 * 1655)$	0.75
$out = 780 - 1606 + (max\{1002, 1084\} - (636 + 1664))$	0.74
$out = (1002 - 1661) - ((1607 - 1379)(625 - 1783)) - \max\{625, 636\}$	0.73
$out = -(\overline{\{-1001, 1270 + 1741\}} + \overline{\{-999, 1606\}})$	0.72
$out = -(-\overline{\{1002, 999\}} + \overline{\{-1473, 1173\}}1473 * (1599 + 723))$	0.72
$out = (636 * 1661 - max\{878, 893\})(636 * 1661 - (1002 - 833))$	0.72
$\begin{array}{l} {\rm out} = -\{{\rm sub}_1, \min\{833, \overline{\{1463, 636\}}\}, {\rm sub}_1, 1603+1661\} \\ {\rm sub}_1 = -\max\{1685, 1002\} \end{array}$	0.71
out = (1001 - 713) + (1715 * 1751 - 636)	0.71
$ \begin{array}{l} \operatorname{out} = \overline{\{(1439 + \overline{\{1533, 1759\}})^2, \operatorname{sub}_1 - \operatorname{sub}_2\}} + (\operatorname{sub}_1 - \operatorname{sub}_2) \\ \operatorname{sub}_1 = \underline{1603 + 1757} + (833 - 1001) \\ \operatorname{sub}_2 = \overline{\{1566, 964\}} \end{array} $	0.71

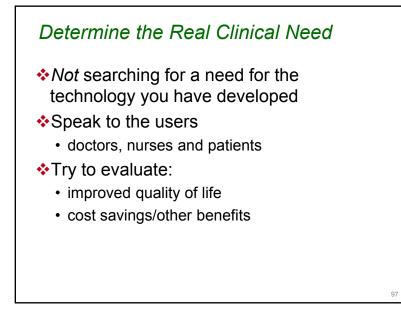












Potential for Commercialisation

Consider commercialiation at the outset

- · establish what is needed and will be useful
- gets the technology to the patient
- raises funding for development
 and further research
- The process itself can have value

Identifying Clinical Collaborators

Essential for:

• data

- establishing the justification of the work
- · obtaining sufficient patient numbers
- obtaining clinical acceptance
- provide clinical interpretation of results

Designing a Clinical Protocol

- Should be clinically led
 - · beware of mission creep/other clinical interests
 - needs to establish ground truth through conventional clinical tests
 - consider engaging a clinical trial unit
- Principal and secondary research questions

100

- Participant sample size
 - with statistical justification
- Inclusion/exclusion criteria
- End points

Datasets

- If using existing data determine its:
 - suitability
 - integrity
- If no existing datasets available
 - · need to acquire patient measurement
 - · ethical approval required
 - · sensors may also require regulatory approval

Applying for Ethical Approval

- Scientific justification
- Clinical protocol
- Participant recruitment procedures
 - participant information sheetinformed consent
- Research may need to undergo external review

Protect Intellectual Property

- Not as difficult as it sounds!
 - initial filing
 - PCT
 - International Phase
- Engage a good patent attorney
- Essential for getting attention
 - · and securing further funding

Regulatory Compliance

- Consider then need for regulatory compliance
- Sometimes required to undertake clinical studies
 - required before sales can begin
 - · often based on self-certification
- Different regulatory bodies for different countries

104

• all based on same ISO standards

103

Undertake Preliminary Clinical Studies Small proof of concept study validates the technology helps determine the clinical need highlights problems/modifications gains supporters/champions

Analysing Data

- Choose an appropriate algorithm
- Consider adopting clinically accepted measures
- Use k-fold cross-validation when dealing with small patient numbers
- Ensembles can be effective in increasing classification performance

Avoid Common Pitfalls

Don't touch the patients (or their data)!

- until you have the necessary ethical approval/consent
- · and possibly other statutory authority
- Be careful what you offer
 - work takes time and consultants can be demanding!
- Expect to do all the running
 - medics are busy
 - many not be used to applying for grants
 - or writing long papers

<section-header><section-header><section-header><section-header><section-header><complex-block><image>

107

Useful Resources

✤Journals

- IEEE Transactions on Evolutionary Computation
- Genetic Programming and Evolvable
 Machines
- IEEE Transactions on Neural Systems and Rehabilitation Engineering
- BioSystems
- International Journal of Computers in Healthcare

109

Useful Resources

Conferences

- IEEE Symposium Series on Computational Intelligence (SSCI) Workshop on Computational Intelligence in Medical Imaging (CIMI)
- Genetic and Evolutionary Computation Conference (GECCO) Medical Applications of Genetic and Evolutionary Computation (MedGEC)

Possible Benefits

- More publications
 - in engineering and medical journals
- More funding sources
 - government grants
 - basic & applied research and knowledge transfer
 - · charities
 - patents and spin-out companies!
- ♦More fun!

Summary

- Medical applications of EC rely on:
 - good collaborations
 - good data
 - appropriate evolutionary algorithms

112

References

- Leonardo Bocchi. 2003. Multiple genetic snakes for bone segmentation. In Proceedings of the 2003 international conference on Applications of evolutionary computing (EvolVorkshops'03), Stefano Cagnoni, Colin G. Johnson, Elena Marchiori, David W. Corne, and Jens Gottlieb (Eds.). Springer-Verlag, Berlin, Heidelberg, 346-356.
- Gabriel Mañana, Fabio González, and Eduardo Romero. 2005. Distributed genetic algorithm for subtraction radiography. In Proceedings of the 2005 workshops on Genetic and evolutionary computation (GECCO '05). ACM, New York, NY, USA, 140-146. DOI=10.1145/1102256.1102288 http://doi.acm.org/10.1145/1102256.1102288
- Ghannad-Rezaie, M., Soltanain-Zadeh, H., Siadat, M.-R. and Elisevich, K.V. 2006. Medical data mining using particle swarm optimization for temporal lobe epilepsy. IEEE Congress on Evolutionary Computation (CEC) 2006, pp. 761–768.
- Alderliesten, T., Bosman, P.A.N., Niessen, W.J. 2007. "Towards a Real-Time Minimally-Invasive Vascular Intervention Simulation System," Medical Imaging, IEEE Transactions on, vol. 26, no. 1, pp. 128–132, doi: 10.1109/TML 2006.886814 URL:
- http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4039522&isnumber=4039521
- Koutsojannis, C. and Hatzilygeroudis, I. 2006. Fuzzy-evolutionary synergism in an intelligent medical diagnosis system. Knowledge-Based Intelligent Information and Engineering Systems. 10th International Conference, KES 2006. Proceedings, Part II. Lecture Notes in Artificial Intelligence, Vol. 4252. Springer-Verlag, pp. 1305– 1312.
- D. Hope, S. Smith, and E. Munday, "Evolutionary algorithms in the classification of mammograms," in IEEE Symposium on Computational Intelligence in Image and Signal Processing, pp. 258–265, 2007.
- J. Walker, K. Völk, S. Smith, and J. Miller, "Parallel evolution using multi-chromosome Cartesian genetic programming," Genetic Programming and Evolvable Machines, vol. 10, no. 4, 2009.
- M. A. Lones, S. L. Smith, A. T. Harris, A. S. High, S. E. Fisher, D. A. Smith and J. Kirkham, "Discriminating Normal and Cancerous Thyroid Cell Lines using Implicit Context Representation Cartesian Genetic Programming," Proc. World Congress on Computational Intelligence (WCCI) 2010, Barcelona, IEEE Press, 2010.

