GECCO 2014

Medical Applications of Evolutionary Computation

Stephen Smith
University of York
York, UK

stephen.smith@york.ac.uk

http://www.sigevo.org/gecco-2014/

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2662-9/14/07.
http://dx.doi.org/10.1145/2598394.2605364

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 senior lecturer 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 ninth year. He is also guest editor for a special issue of Genetic Programming and Evolvable Machines (Springer) on medical applications and co-editor 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.

The City of York

The 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

Overview

- Introduction
- Previous work
- Why consider a medical application?
- Choosing a Medical Application
- Data sets
- Choice of evolutionary algorithm
- Data input, output and fitness function
- Case examples
- Getting started

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 Knowledge-based Systems
- Clinical Diagnosis and Therapy

Medical Imaging and Signal Processing

Ballerini and Bocchi [1]

- Automated image segmentation
- Bone age assessment
  - to detect abnormality in skeletal growth
- Multiple genetic snakes
  - planar active contours optimised by GAs
  - elastic force connects adjacent snakes

Source: Ballerini and Bocchi [1]

Medical Imaging and Signal Processing

Mañana, González and Romero [2]

- Image Registration
- Intra-oral radiographs
- Distributed adaptive GA
  - to find an optimal affine
- Genome
  - 4 floating point numbers
    - scale factor
    - rotational angle
    - X &Y translations

Source: Mañana, González and Romero [2]

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)

Source: Ghannad-Rezaie *et al.* [3]

Modelling and Simulation of Medical Processes

Bosman and Alderliesten [4]

- Medical simulation
  - for minimally invasive vascular intervention
  - used to train clinicians
- Minimisation of energy
  - associated with trajectory of guide wire
  - iterative density estimation evolutionary
  - accurate and closer to real-time simulation of the intervention

Source: Bosman and Alderliesten [4]

Clinical Expert Systems and Knowledge-based Systems

Koutsojannis and Hatzilygeroudis [5]

- Diagnosis and treatment of blood gas disturbances
  - observed in intensive care unit patients
- Fuzzy expert system incorporating an EA
  - differential EA fine-tunes membership functions of the fuzzy variables

Source: Koutsojannis and Hatzilygeroudis [5]

Why Consider a Medical Application?

- Interesting
  - real problems
  - potential to help real people
  - cross-disciplinary collaborations
- Good source of funding
  - government
  - charities
  - commercialization opportunities
- Good for publications
  - exploit both engineering and medical literatures

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”

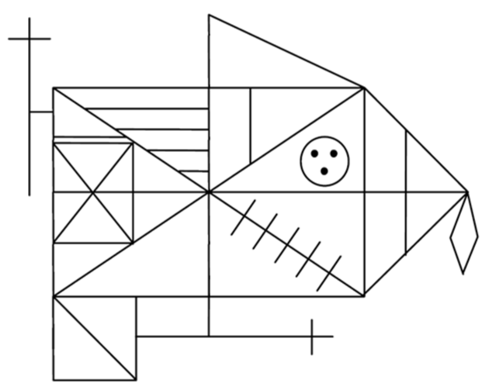
Datasets

- 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 cross-validation

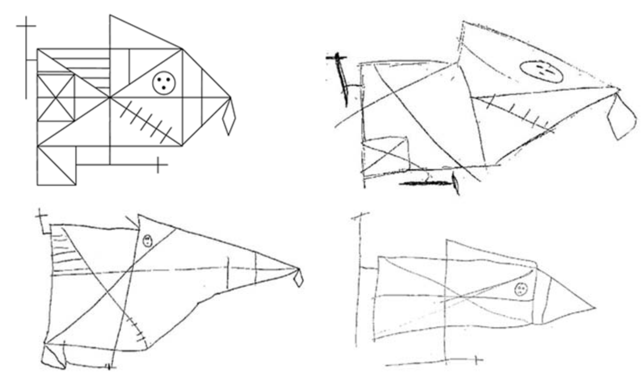
Datasets

- 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
    - to sample representative demographic population
- Engage a health statistician ***before*** you start

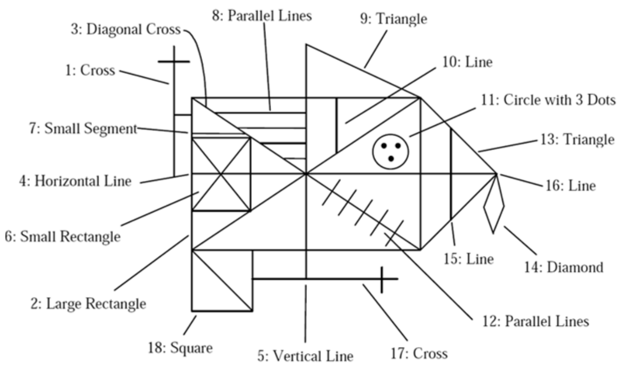
Complex Figure Copying Test



Example Patient Drawings



Subjective Mark Scheme

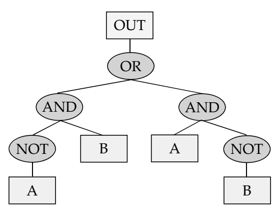


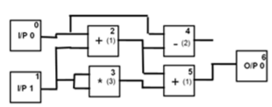
Choice of Evolutionary Algorithm

- Many to choose from
- Choice often based on:
  - previous experience
  - match to data representation
  - performance in previous similar applications
- Consider conventional statistical analysis
  - comparison of performance
  - justification for the use of an EA
  - understand the nature of the data
- Ensembles of EAs have potential

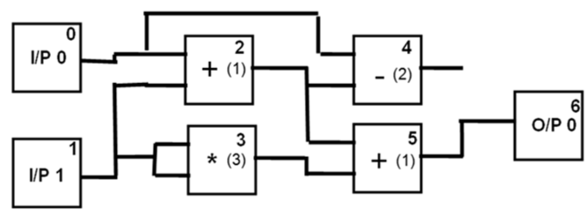
Cartesian Genetic Programming (CGP)

- Form of genetic programming
  - originators Miller and Thomson (1997)
- Genotype is a list of integers
  - that represent the program elements
  - how they are connected together
- Representation a directed graph
  - simple, flexible and convenient for many problems

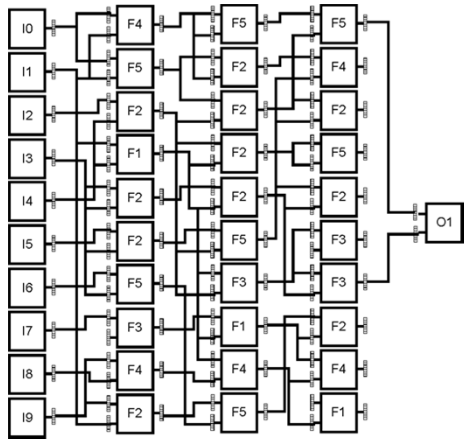




Example CGP Network

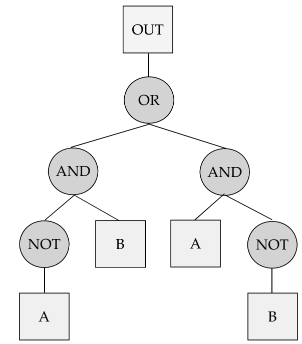


0 1 1   1 1 3   0 2 2   2 3 1   5

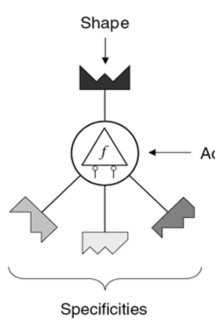


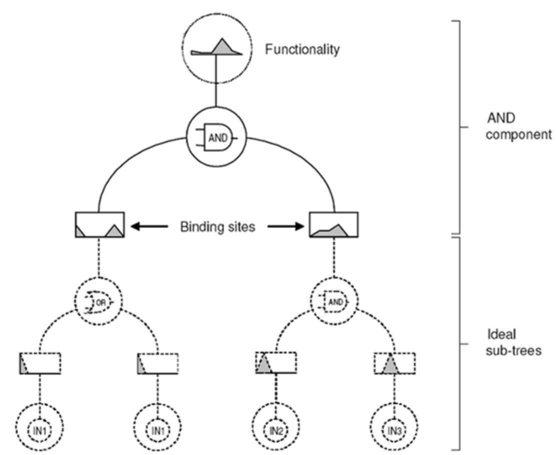
Implicit Context Representation

- Consider standard GP
- Parse tree structure
- Mutation and crossover



Enzyme Genetic Programming (Lones and Tyrrell)

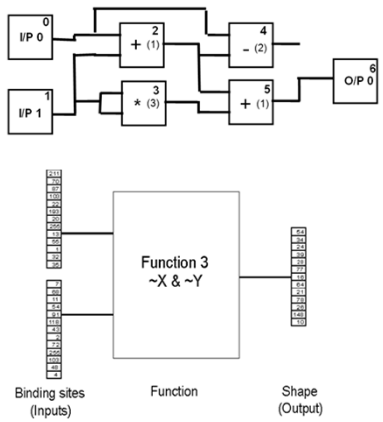




Implicit Context Representation for CGP

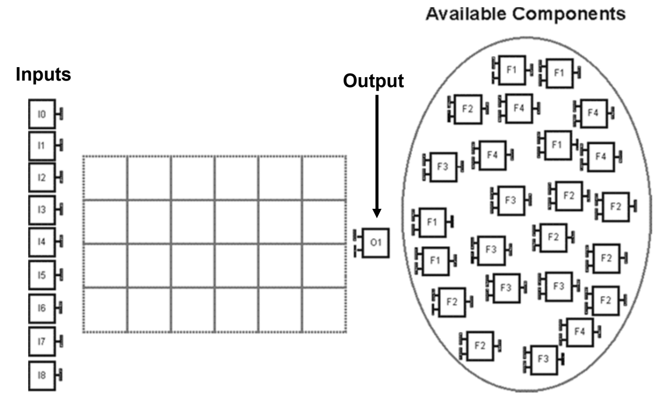
Mutation of components'

- Binding sites' shapes
- Function



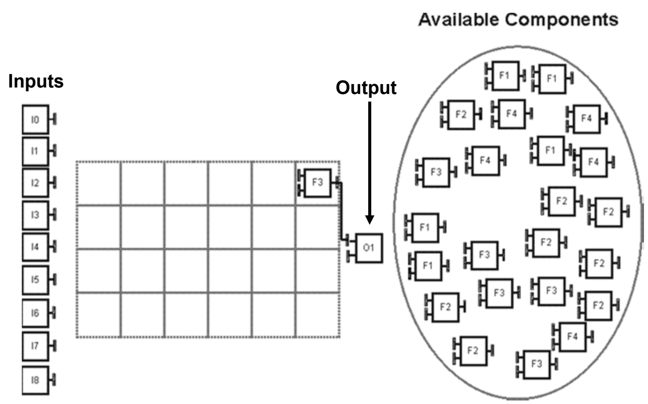
THE UNIVERSITY of York

Formation of Network



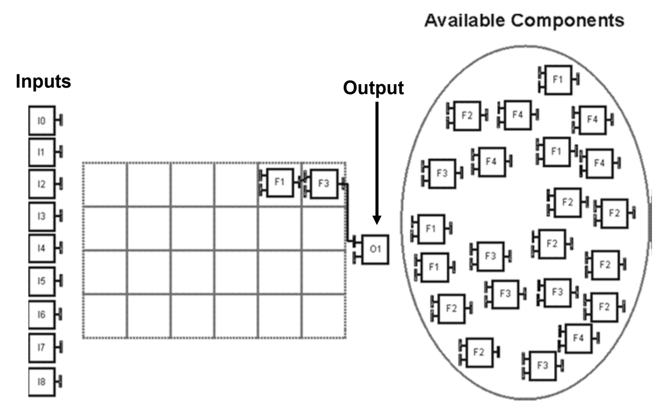
THE UNIVERSITY of York

Formation of Network



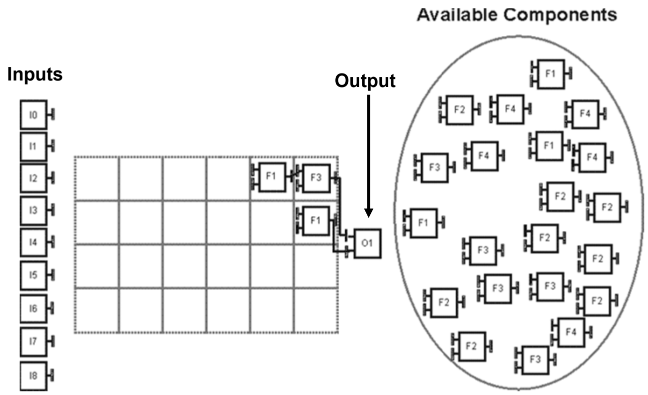
THE UNIVERSITY of York

Formation of Network

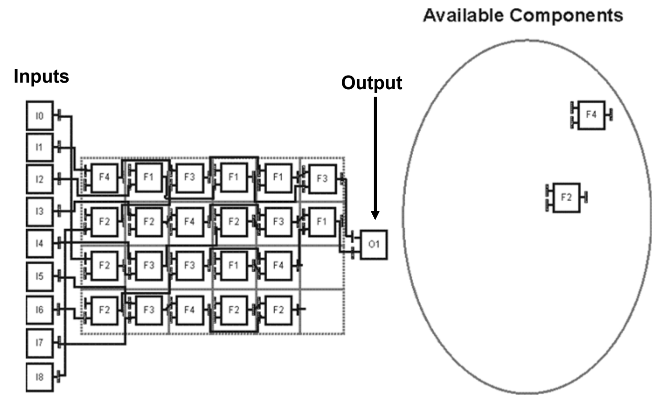


THE UNIVERSITY of York

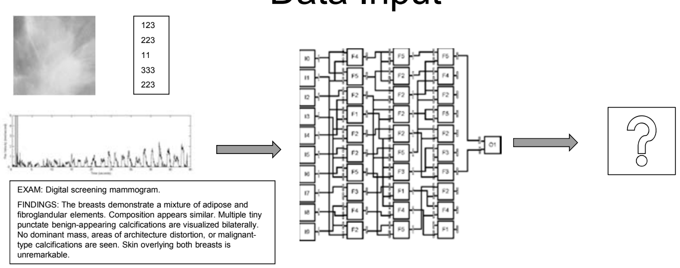
Formation of Network



Formation of Network

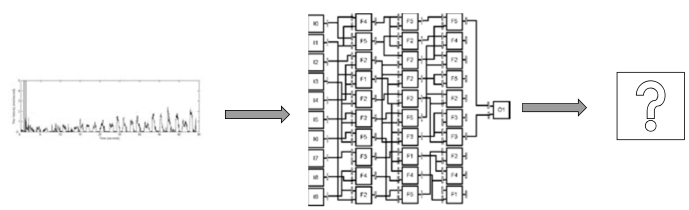


Data Input



- Many different possible modalities, e.g.
  - image
  - continuous and discrete measurements
  - text

Desired Output



- Often a two class problem – disease/no disease
  - Sometimes multiclass (typically 5)
- Requires thresholding of output or other approaches

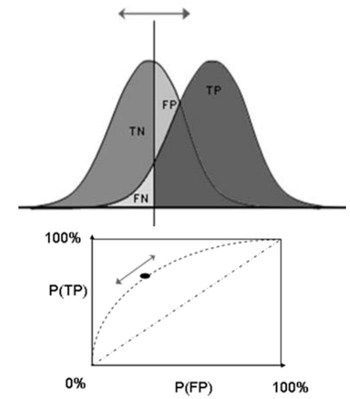
Fitness Function

- In clinical studies often based on sensitivity and specificity
  - sensitivity = TP/P
  - specificity = TN/N

  Where:
  TP number of true positive instances correctly classified
  TN number of true negative instances correctly classified
  P is number of positive instances that truly exist
  N is number of negative instances that truly exist

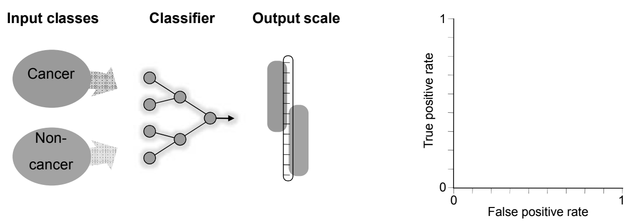
Choice of Threshold Value

- Can be used to weight classification
- According to the cost associated with a misclassification
- Often summarised using ROC curves



ROC Analysis

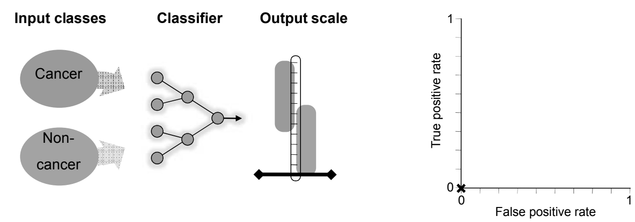
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

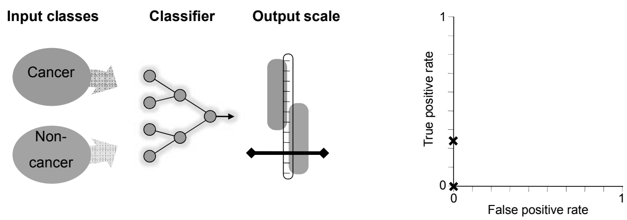
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

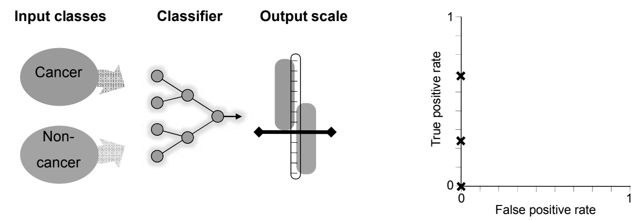
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

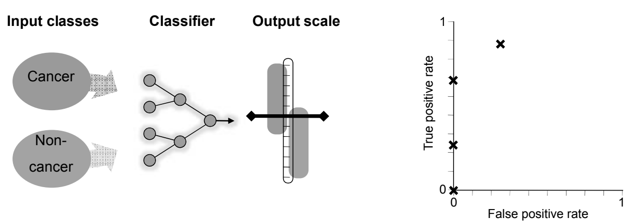
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

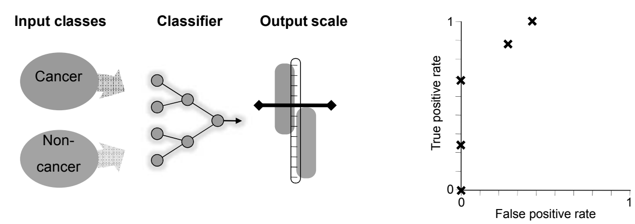
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

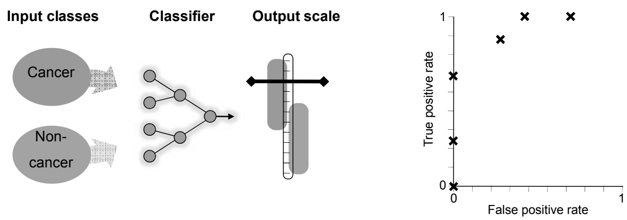
- Receiver Operating Characteristic (ROC) curves:



Slide: Mic Lones

ROC Analysis

- Receiver Operating Characteristic (ROC) curves:

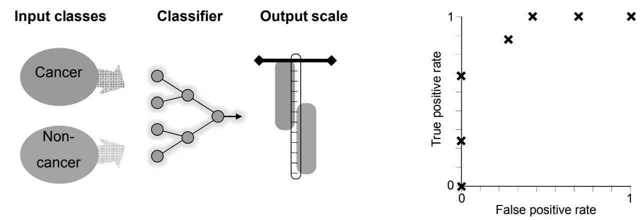


Slide: Mic Lones

THE UNIVERSITY of York 45

ROC Analysis

- Receiver Operating Characteristic (ROC) curves:

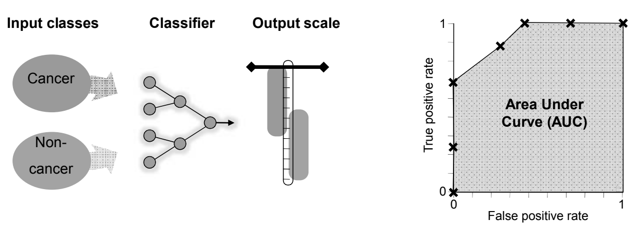


Slide: Mic Lones

THE UNIVERSITY of York 46

ROC Analysis

- Receiver Operating Characteristic (ROC) curves:

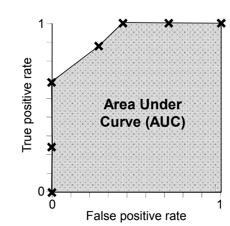


Slide: Mic Lones

THE UNIVERSITY of York 47

ROC Analysis

- Receiver Operating Characteristic (ROC) curves:
  - Area Under Curve (AUC)
    - Probability of ranking a class member higher than a non-member
    - *i.e.* p[C(cancer)>C(non-cancer)]

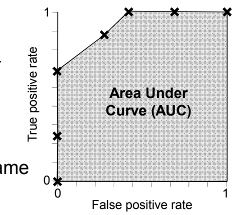


Slide: Mic Lones

THE UNIVERSITY of York 48

ROC Analysis

- Receiver Operating Characteristic (ROC) curves:
  - Area Under Curve (AUC)
    - Probability of ranking a class member higher than a non-member
    - *i.e.* p[C(cancer)>C(non-cancer)]
    - AUC of 0.5 equivalent to random classification
    - A classifier with AUC = *n* has the same power as one with AUC = 1-*n*
    - AUCs recently attracted criticism



Slide: Mic Lones

ROC Analysis

- Multi-class ROC analysis:
  - Hand and Till, 2001:

$$\mathrm{AUC}_{\text{multiclass}} = \frac{2}{|C|(|C|-1)} \sum_{\{c_i,c_j\}\in C} \mathrm{AUC}(c_i, c_j)$$

  - ≈ 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.

Case Example 1: Breast Cancer

(www.breastcanceruk.org.uk/statistics/2004-6)

- 207,000 women diagnosed in the US in 2010
  - 40,000 died as a result of the disease
- 1 in 9 women will suffer from breast cancer at some time in their life
- Most commonly diagnosed cancer in women under the age of 35
- Accounts for some 25% of all cancer cases in women worldwide

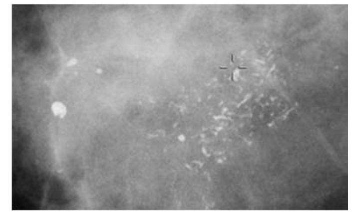
Mammograms

- Assessed subjectively
- Size, shape and texture of
  - microcalcifications
  - spiculated masses

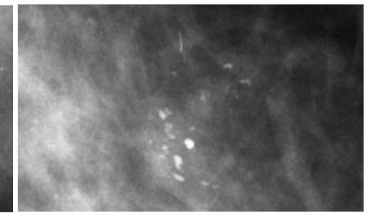
Microcalcifications

- Calcium deposits
  - secretions from ductal structures
- Occur in clusters
- 40-50% cancerous
- Discriminated by:
  - shape
  - size
  - texture
  - distribution

Microcalcifications



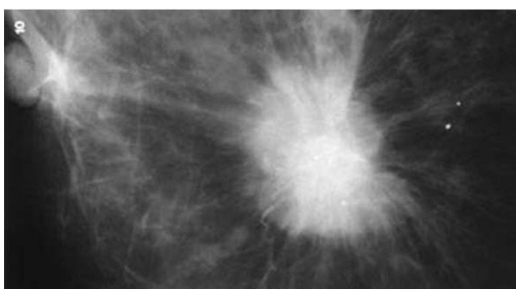
Example cluster of malignant microcalcifications



Example cluster of benign microcalcifications

(Images courtesy of the *Radiology Assistant:* www.radiologyassistant.nl)

Spiculated Masses



Example spiculated mass

(Image courtesy of the *Radiology Assistant:* www.radiologyassistant.nl)

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
  - digitization
  - resolution
  - file format

THE UNIVERSITY of York 57

Using Existing Datasets

- Using other peoples datasets may have implications!
- LLNL/UCSF database
- From "Somewhere in England"

THE UNIVERSITY of York 58

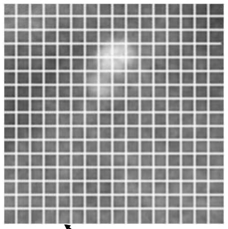
Constructing a New Dataset

- Requires:
  - clinical cooperation
  - authority
  - suitable software tools
    - especially, the user interface

THE UNIVERSITY of York 59

Regions of Interest (ROI)

- Regions showing microcalcifications are identified (128x128 pixel)
- These are divided into smaller sections called 'parts' (8x8 pixels)



8x8 pixels

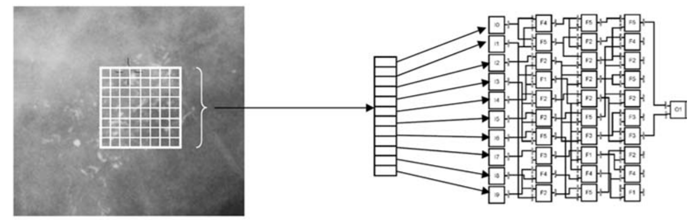
THE UNIVERSITY of York

Application of CGP

- CGP as an optimizer
- CGP as a feature extractor and classifier

CGP as an Optimizer

Hope, Smith and Munday [6]



Region of interest in mammogram

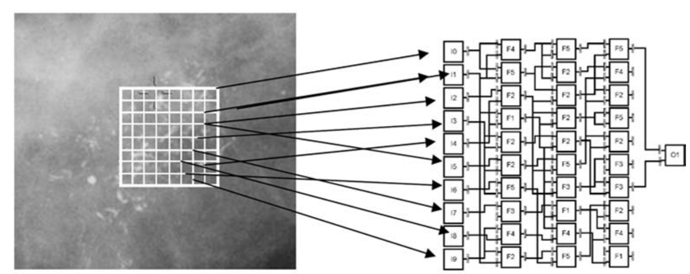
SGLD Matrix

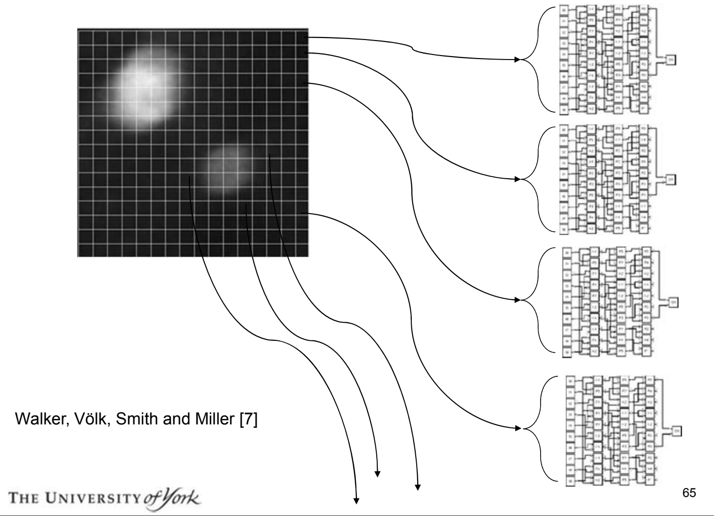
CGP network

Spatial Gray Level Dependence Matrix

- Also known as spatial gray co-occurrence matrix
- Considers positional information of pixels
- Following features from the SGLD matrices
  - at 45°, 90°, 135° and 180°
  - maximum value
  - second moment
  - third moment

CGP as Feature Extractor and Classifier

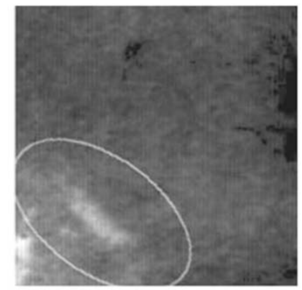




Walker, Völk, Smith and Miller [7]

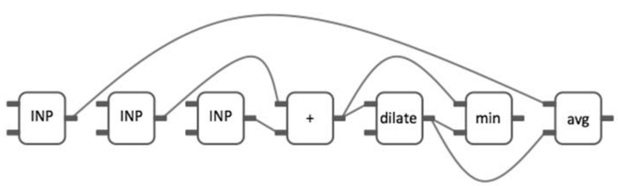
Example Results

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0



Additional Approaches

- Use OpenCV image functions as CGP function set (Simon Harding, CGP-IP)



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

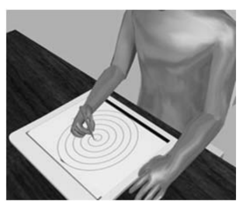
- Tremor
  - 3Hz-8Hz
- Rigidity
  - cogwheel
- Bradykinesia
  - slowing in movement
- Postural instability
  - poor balance

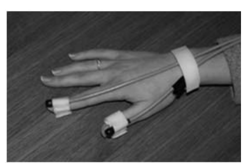
Conventional Diagnosis of Parkinson's Disease

- 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

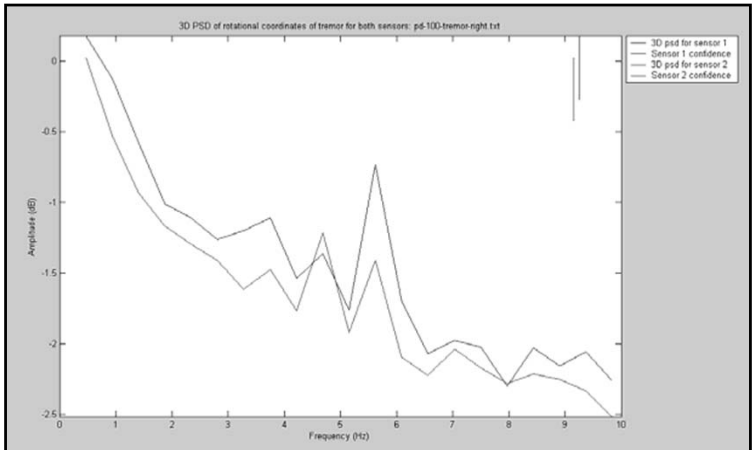
- Based on neuropsychological tasks
  - figure copying tasks
  - finger tapping tasks
  - measurement of tremor at rest
- Patient's movements digitised in real-time





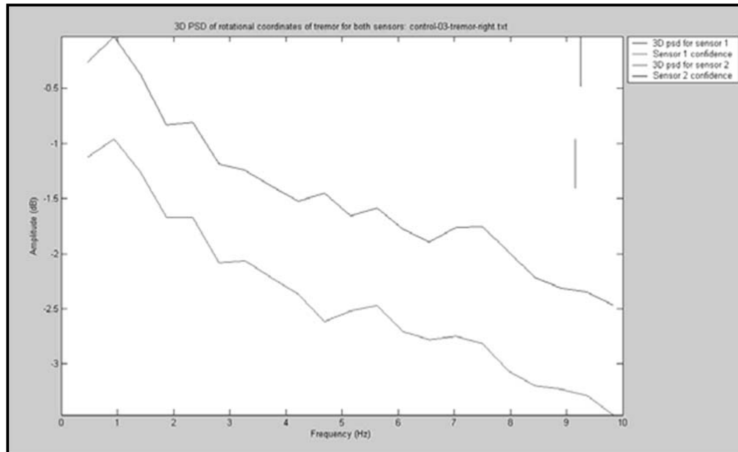
Data Analysis

- Initially, results poor
- Closer inspection of patient data
  - patients with Parkinson's disease
  - controls with no known neurodegenerative condition



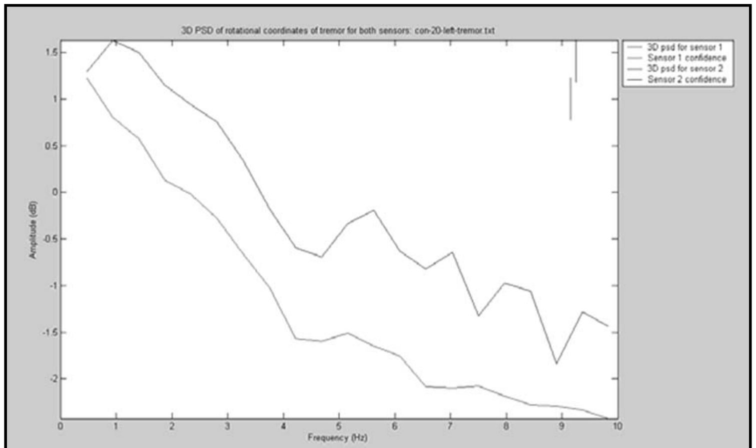
Typical Parkinson’s Disease Patient FFT

THE UNIVERSITY of York 73



Typical Control FFT

THE UNIVERSITY of York 74

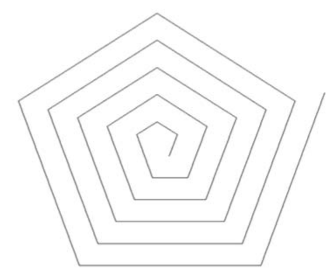


Problematic Control FFT

THE UNIVERSITY of York 75

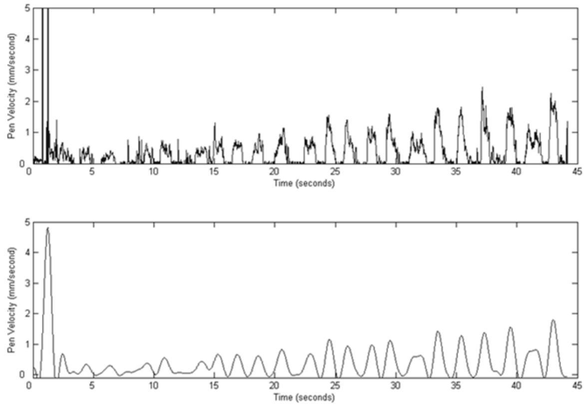
Figure-Copying Tasks

- Choice of task domain important
- Used to emphasize symptoms under consideration



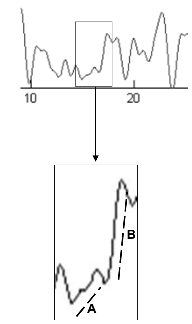
THE UNIVERSITY of York 76

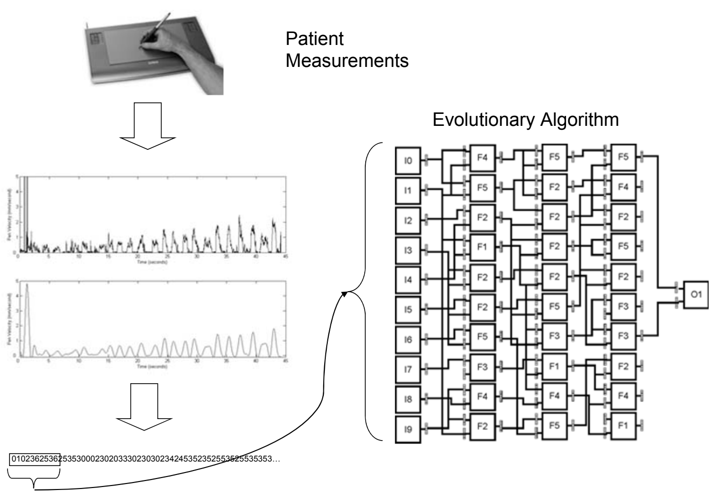
Velocity Profile of Patient's Response



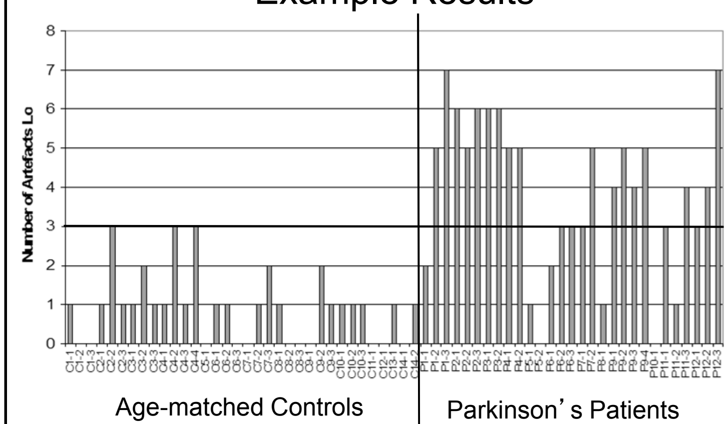
Identification of Artefacts

- Artefacts unique to Parkinson's patients
  - determined through visual inspection
- Indicates a two-stage acceleration
- Consistent with hesitation
  - Bradykinesia?

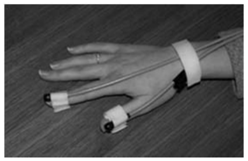




Example Results

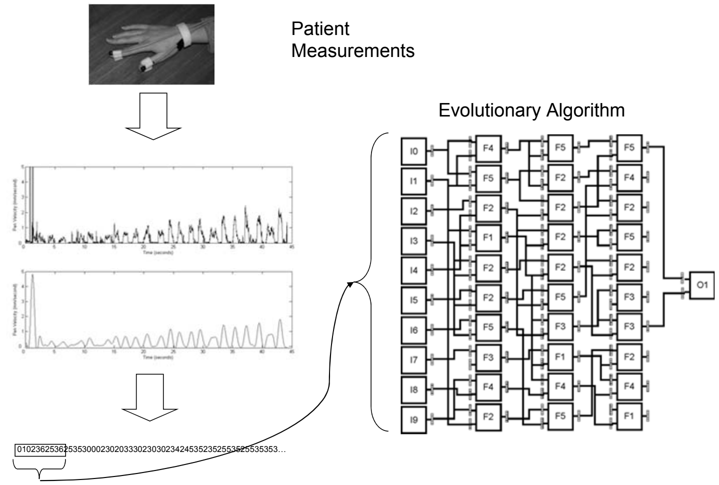


Finger Tapping Task



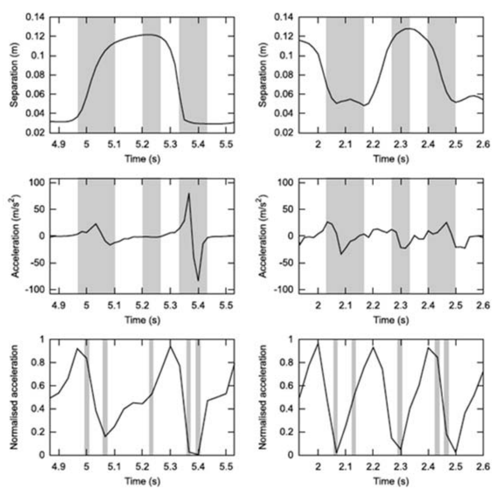
THE UNIVERSITY of York

81



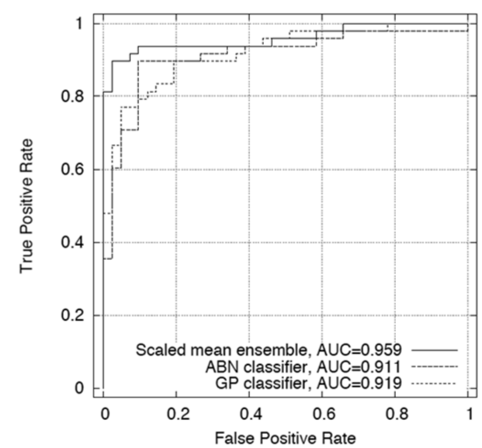
THE UNIVERSITY of York

82



THE UNIVERSITY of York

83



THE UNIVERSITY of York

84

Case Example 3: Thyroid cancer

- Incidence of 1 in 842 for men, 1 in 324 for women[1]
- 4 types, in order of increasing aggressiveness:
  - **Papillary carcinoma**
    - 70-80% of cases, low mortality rate
  - **Follicular carcinoma**
    - 10-15%, fairly low mortality rate
  - **Medullary carcinoma**
    - 5-10%, moderate mortality rate
  - **Anaplastic carcinoma**
    - <5%, high mortality rate

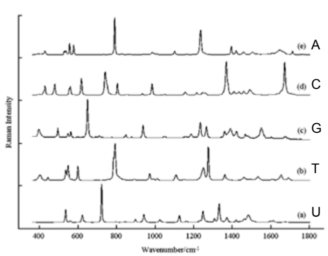
[1] UK Thyroid Cancer Statistics, Cancer Research UK, 2008.

Raman spectroscopy

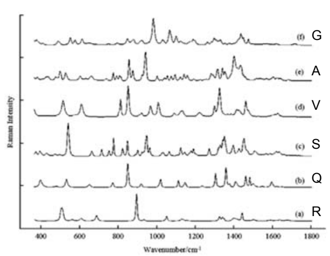
- Optical methodology
  - Uses monochromatic light
  - Measures energy states
- Advantages
  - Detailed
  - Fast
  - Non-invasive
  - Non-destructive
  - Uses small samples
- Disadvantages
  - Complex signals
  - Difficult to analyse

Biochemical spectra

DNA and RNA bases

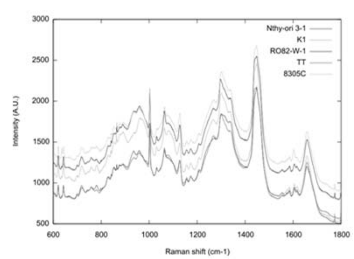


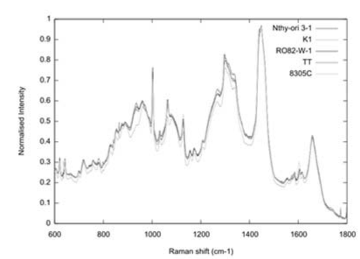
Amino acids



J. De Gelder, K. De Gussem, P. Vandenabeele and Luc Moens, Reference Database of Raman spectra of biological molecules, *Journal of Raman Spectroscopy* 38:1133-1147, 2007.

Thyroid cell lines

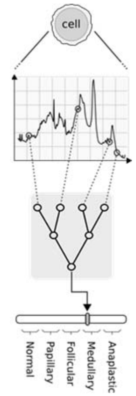




Class	Name	Cell Type	Samples
1	Nthy-ori 3-1	Normal follicular epithelial	30
2	K1	Papillary thyroid carcinoma	25
3	RO82-W-1	Follicular thyroid carcinoma	25
4	TT	Medullary thyroid carcinoma	24
5	8305C	Anaplastic thyroid carcinoma	25

Methodology

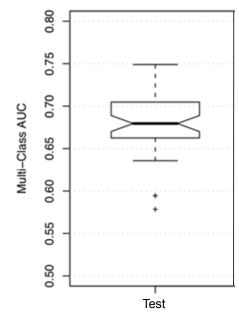
- Evolve classifiers
  - Terminal set: Raman offsets {600,…,1800}
  - Function set: +, -, *, mean, min, max, negate
- Using implicit context representation CGP [Smith, 2005]
  - Cartesian Genetic Programming [Miller and Thomson, 2000]
  - Implicit context representation [Lones and Tyrrell, 2002]
- Fitness is the ability to separate cancer classes
  - In order of biological aggressiveness
  - Measured using multi-class area under ROC curve



S. Smith et al. An implicit context representation for evolving image processing filters, *EvoIASP 2005.*
J. Miller and P. Thomson, Cartesian genetic programming, *EuroGP 2000.*
M. Lones and A. Tyrrell, *Genetic Programming and Evolvable Machines* 3(2):193-217.

Results

All runs

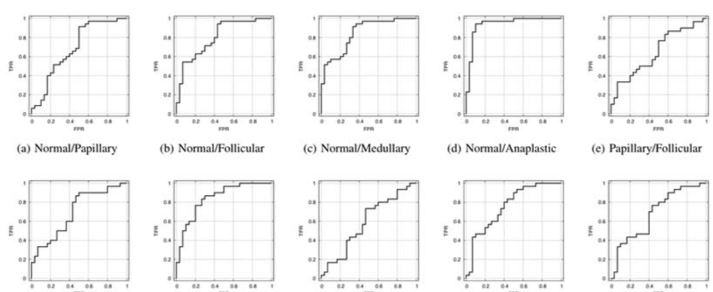


Best run

Classes	Train	Validate	Test
1/2/3/4/5	**0.75**	**0.78**	**0.75**
1/2	0.69	0.64	0.61
1/3	**0.80**	**0.83**	0.74
1/4	**0.85**	**0.94**	**0.88**
1/5	**0.94**	**0.98**	**0.91**
2/3	0.65	0.69	0.62
2/4	0.70	**0.82**	**0.79**
2/5	**0.84**	**0.89**	**0.83**
3/4	0.59	0.65	0.71
3/5	**0.76**	**0.76**	**0.78**
4/5	0.68	0.62	0.60

Results

ROC curves for highest scoring classifier



(a) Normal/Papillary (b) Normal/Follicular (c) Normal/Medullary (d) Normal/Anaplastic (e) Papillary/Follicular
(f) Papillary/Medullary (g) Papillary/Anaplastic (h) Follicular/Medullary (i) Follicular/Anaplastic (j) Medullary/Anaplastic

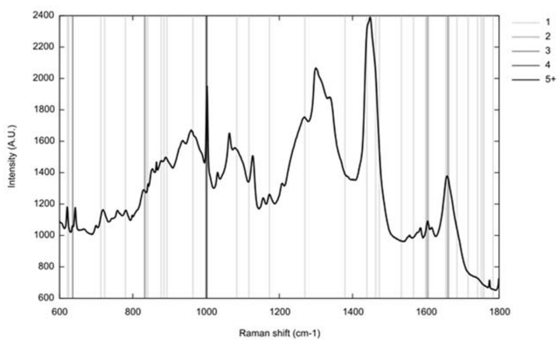
Results

The 10 highest scoring classifiers

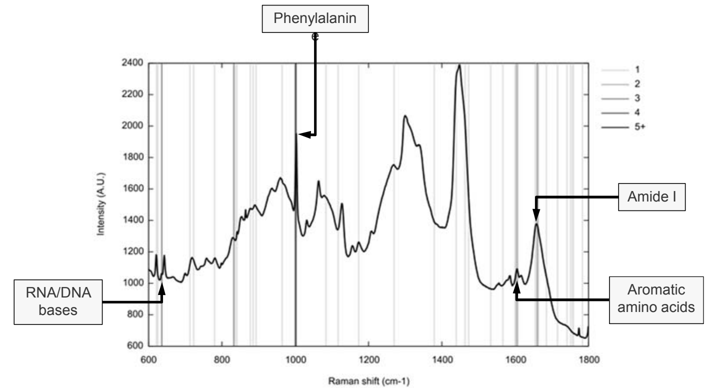
Expression	Test AUC
$\mathrm{out} = \overline{\{885, 1001\}} - 1607 - \max\{1656, 1117\}$	0.75
$\mathrm{out} = (841 + 1002) - (836 + 1600) + (\max\{-622, 1657 * 1655\} - 1657 * 1655)$	0.75
$\mathrm{out} = 780 - 1606 + (\max\{1002, 1084\} - (636 + 1664))$	0.74
$\mathrm{out} = (1002 - 1661) - ((1607 -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

Results

Raman shifts referenced by the 10 highest scoring classifiers



Results

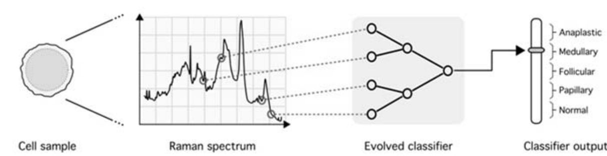


Experimental settings

- 50 runs, 50 generation limit, population of 500
- CGP grid size of 5x5
- Using training, validation and test sets

Function	Description
x + y	Returns the sum of its inputs
x – y	Returns the difference of its inputs
x * y	Returns the product of its inputs
$\overline{\{x,y\}}$	Returns the mean of its inputs
min{x,y}	Returns the lesser of its inputs
max{x,y}	Returns the greater of its inputs
-x	Returns its input multiplied by -1

Overview



Planning your own study

- Determine the *real* clinical need
  - and the potential for commercilasation
- Identify clinical collaborators
- Design a clinical protocol
- Protect intellectual property
- Apply for ethical approval
- Obtain regulatory compliance
- Undertake clinical studies
- Processing data

THE UNIVERSITY of York 97

Determine the *Real* Clinical Need

- *Not* searching for a need for the technology you have developed
- Speak to the users
  - doctors, nurses and patients
- Try to evaluate:
  - improved quality of life
  - cost savings/other benefits

THE UNIVERSITY of York 98

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

THE UNIVERSITY of York 99

Identifying Clinical Collaborators

- Essential for:
  - data
  - establishing the justification of the work
  - obtaining sufficient patient numbers
  - obtaining clinical acceptance
  - provide clinical interpretation of results

THE UNIVERSITY of York 100

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

THE UNIVERSITY of York

101

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

THE UNIVERSITY of York

102

Applying for Ethical Approval

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

THE UNIVERSITY of York

103

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

THE UNIVERSITY of York

104

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
  - all based on same ISO standards

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

Useful Resources

- Books

THE UNIVERSITY of York 109

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

THE UNIVERSITY of York 110

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)

THE UNIVERSITY of York 111

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!

THE UNIVERSITY of York 112

Summary

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

References

1. Leonardo Bocchi. 2003. Multiple genetic snakes for bone segmentation. In Proceedings of the 2003 international conference on Applications of evolutionary computing (EvoWorkshops'03), Stefano Cagnoni, Colin G. Johnson, Elena Marchiori, David W. Corne, and Jens Gottlieb (Eds.). Springer-Verlag, Berlin, Heidelberg, 346-356.
2. 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
3. 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.
4. 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/TMI. 2006.886814 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4039522&isnumber=4039521
5. 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.
6. 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.
7. 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.

EPSRC Pioneering research and skills
H·Y·M·S THE HULL YORK MEDICAL SCHOOL
PARKINSON'S UK CHANGE ATTITUDES. FIND A CURE. JOIN US.
ROYAL ACADEMY OF ENGINEERING
wellcome trust
ClearSky