Discriminating Normal and Cancerous Thyroid Cell Lines using Implicit Context Representation Cartesian Genetic Programming

Michael A. Lones, *Member, IEEE*, Stephen L. Smith, Andrew T. Harris,
Alec S. High, Sheila E. Fisher, D. Alastair Smith and Jennifer Kirkham

***Abstract*— In this paper, we describe a method for discriminating between thyroid cell lines. Five commercial thyroid cell lines were obtained, ranging from non-cancerous to cancerous varieties. Raman spectroscopy was used to interrogate native cell biochemistry. Following suitable normalisation of the data, implicit context representation Cartesian genetic programming was then used to search for classifiers capable of distinguishing between the spectral fingerprints of the different cell lines. The results are promising, producing comprehensible classifiers whose output values correlate with biological aggressiveness.**

I. Introduction

Papillary cancer is the most common thyroid malignancy, followed by follicular carcinoma. Both of these cancers have a high chance of cure. Medullary thyroid carcinoma, which is associated with a genetic predisposition, occurs less frequently, but has a higher mortality rate. Anaplastic carcinoma is the rarest variety, usually occurring in the elderly with a very poor prognosis. Diagnosis of thyroid cancer is made on needle biopsy. Should this not prove adequate, removal of half of the gland is required with histological analysis. This is an invasive procedure usually taking two to three weeks before results are confirmed.

Recently, there has been growing interest in the use of optical methodologies to identify diseased cells and tissues, since these have the potential to provide a non-invasive, rapid and objective diagnosis prior to the onset of visible symptoms [6]. In this paper, we focus on the use of Raman spectroscopy, an optical methodology which provides considerable information about the chemical composition of a sample, making it of particular interest for diagnosing diseases, such as cancer, which have a complex biochemical signature [11].

A Raman spectrum is generated when a sample is illuminated by a monochromatic light, causing photons to be absorbed and reemitted at different (or *shifted*) frequencies determined by the energy states of its constituent molecules. Raman spectroscopy provides a highly detailed biochemical fingerprint of the target material. However, biological cells are biochemically complex, making interpretation of Raman spectra highly challenging. Chemical components of the cell will each produce peaks in the spectrum, with those of the highest quantity producing the largest peaks. Thus this method is both qualitative and quantitative in nature. Interpreting this biochemical information requires sophisticated data analysis methods capable of discriminating between subtle differences in the data sets.

In this paper, we describe an approach to discriminating between the Raman spectra of normal and cancerous thyroid cell lines using implicit context representation Cartesian genetic programming (IRCGP), a form of genetic programming which has previously been used to successfully learn biomedical classifiers [8, 18, 20]. In particular, we take a multi-class approach, training individual classifiers to distinguish between five thyroid cell lines (one normal and four cancerous) ordered by increasing biological aggressiveness.

The paper is organised as follows: In Section II, we review previous and related work. Section III describes materials and methods. Section IV provides experimental results and analysis. Section V concludes.

II. Previous and Related Work

Spectral analysis using evolutionary algorithms: Several types of evolutionary algorithm have been used to interpret Raman spectra. In [12], the authors used a genetic algorithm (GA) to identify the Raman spectral components associated with different types of plastic. In [2], the authors used a hybrid GA-immune algorithm to identify chemicals present in composite spectra. In [9], the authors used standard genetic programming (GP) to identify solvent samples from their Raman spectra, showing the approach to be superior to those based upon neural networks and more traditional multivariate data analysis techniques. GP has also been used to interpret Fourier Transform Infrared (FTIR) spectral data, showing good explanatory power when compared to more traditional analysis techniques [4].

Classification using IRCGP: We have previously looked at the feasibility of using IRCGP to discriminate Raman spectra, showing that it is capable of generating classifiers able to discriminate blood samples from subjects with and without head and neck cancers [8]. We have also shown how this IRCGP-based method (in the context of measuring visuo-

Michael Lones and Stephen Smith are with the Intelligent Systems Research Group, Department of Electronics, University of York, Heslington, York, UK, YO10 5DD. Email: {mal503,sls5}@ohm.york.ac.uk. Web: http://www.elec.york.ac.uk/research/intSys.html.

Andrew Harris and Jennifer Kirkham are with the Department of Oral Biology, and Alec High is with the Department of Pathology, Leeds Dental Institute, University of Leeds, Leeds, UK. Email: andrew.harris@doctors.org.uk, J.Kirkham@leeds.ac.uk, and A.S.High@leeds.ac.uk. Web: http://www.leeds.ac.uk/dental/.

Sheila Fisher is with the Section of Experimental Therapeutics, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, UK, and the School of Health Studies, University of Bradford, Bradford, UK. Email: s.e.fisher@doctors.org.uk.

Alastair Smith is with AVACTA group plc, York Biocentre, York Science Park, York, UK. Email: alastair@avacta.com. Web: http://www.avacta.com.

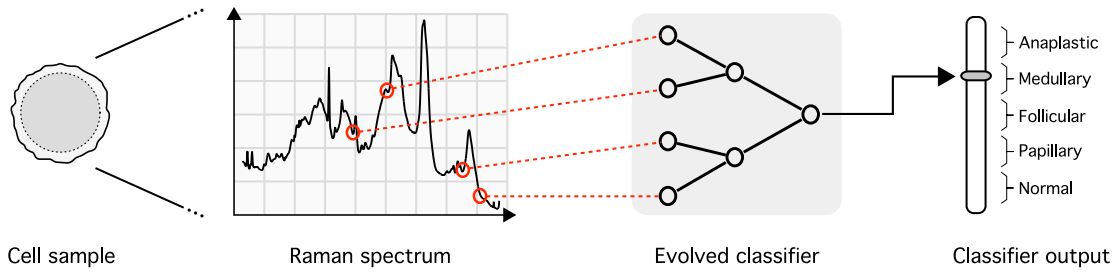


Fig. 1: Classifying the Raman spectrum of a thyroid cell.

spatial ability [18]) can be extended to problems of multi-class discrimination, such as the one addressed in this paper.

Spectral discrimination of thyroid cancer: In [21], the authors used discriminative linear analysis to compare the Raman spectra of benign thyroid tumours and papillary and follicular carcinomas, finding good discrimination between benign and malignant tissues. Our own initial study of two cell lines, taken from normal thyroid tissue and anaplastic carcinoma, supports this conclusion, showing that a self-organising map can accurately distinguish the Raman spectra of these two cell lines [7]. The current paper extends this initial study by applying the multi-class IRCGP approach, developed in [18], to discriminate between five thyroid cell lines: one normal and four cancerous.

III. Materials and Methods

We used IRCGP to evolve classifiers capable of discriminating between Raman spectra of normal thyroid cells and each of the four main classes of thyroid cancer on a single, continuous-valued, output scale (see schematic in Figure 1).

A. Implicit Context Representation CGP

Implicit context representation Cartesian genetic programming (IRCGP) [19] is a graph-based genetic programming system which uses the idea of implicit context [13–15] to provide positional independence to evolving solutions. IRCGP is a variant of Cartesian genetic programming (CGP) [17]. Like CGP, an IRCGP solution consists of an n-dimensional grid (where n is typically 1 or 2) in which each grid location contains a function, and program inputs and outputs are delivered to and taken from specific grid locations. However, unlike standard CGP, interconnections between functions, inputs and outputs are specified in terms of a component's *functionality profile*: a vector describing the component's functional context within the program. Since functionality profiles are independent of grid position, this means that a program's behaviour is more likely to be preserved when variation operators modify a component's absolute or relative grid position. In particular, this has been shown to improve performance when crossover operators are used [1, 15]. More details about the form of IRCGP used in this work can be found in [18].

B. Data Collection and Preprocessing

Collection: Five human thyroid cell lines (see Table I), one normal and four cancerous, were cultured in media. Up to 30 cells were sampled from each cell line, formalin-fixed and placed onto calcium fluoride slides. Exciting light was provided by a cavity laser set at 783 nm, and 5 Raman spectra were collected for each cell sample using a thermoelectrically cooled CCD camera attached to a Renishaw 'System 1000' Raman microscope. Further details of the experimental setup can be found in [7].

Normalisation: When collecting spectral data, variability of experimental conditions, sample consistency, and instrumentation lead to differences in spectral intensity. In Figure 2a, this is particularly evident in differences in the total energies (area under curves) of the mean spectra for each of the cell lines. To compensate for this, each sample's spectrum was linearly scaled to the unit interval, mapping the highest intensity to 1 and the lowest intensity to 0. Figure 2b shows the resulting mean spectra for each cell line.

Data Sets: The normalised data were divided uniformly into training, validation and test sets, maintaining a within-class ratio of 2:1:1. To prevent possible bias, spectra from the same cell sample were kept together.

C. Classifier Evaluation

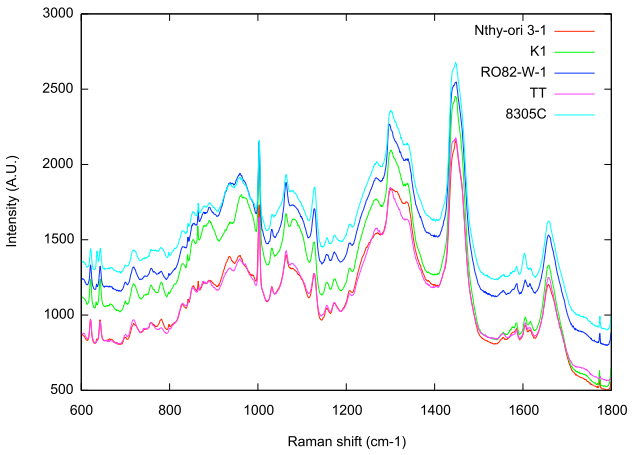
Receiver Operating Characteristic (ROC) analysis is used to measure an evolved classifier's fitness — its ability to discriminate between data classes. A ROC curve plots true positive rate (TPR) against false positive rate (FPR) across the range of possible classification thresholds, where:

$$\text{TPR} = \frac{\text{Number of positive examples correctly classified}}{\text{Number of positive examples}}$$

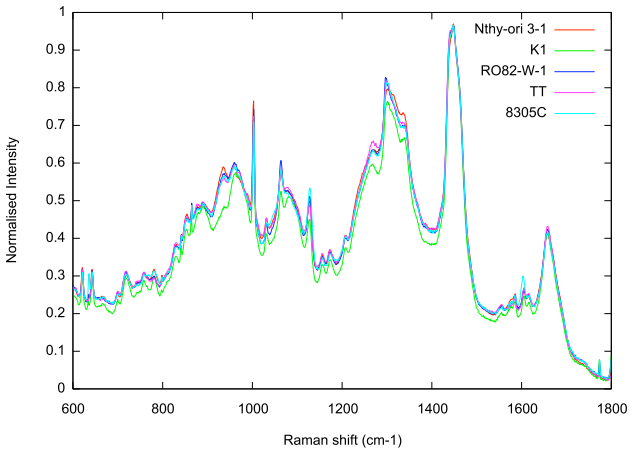
$$\text{FPR} = \frac{\text{Number of negative examples incorrectly classified}}{\text{Number of negative examples}}$$

TABLE I: 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



(a) Original



(b) Normalised

Fig. 2: Mean Raman spectra for each cell line.

The area under a ROC curve (known as AUC) is often used as a measure of classifier accuracy, since it is equivalent to the probability that the classifier will rank a randomly chosen class member higher than a randomly chosen non-member [3]. AUC scores fall within the range $[0, 1]$, where 1 indicates perfect discrimination of class members from non-members, 0.5 indicates no ability to discriminate, and 0 indicates that non-members are always ranked higher than members (i.e. perfect classification can be achieved by inverting the classifier's output).

The AUC metric can be extended to multi-class classifiers by taking the mean of the AUCs between each pair of classes [5], thus measuring the overall pairwise discriminability of the classifier — in effect, how well the classifier separates the classes within its output range. Hand and Till [5] define this metric as:

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

where C is the set of classes and $\text{AUC}(c_i, c_j)$ is the area under the ROC curve when separating classes c_i and c_j.

D. Experimental Settings

We carried out 50 runs of 50 generations using a population of 500 classifiers. All members of the population were evaluated separately upon both the training and validation data sets. At the end of a run, the solution with the highest validation set accuracy (during the whole run) was evaluated upon the test set, giving an unbiased measure of its fitness.

Child solutions were generated using uniform crossover and mutation in equal proportion. The mutation rate (determined experimentally) was 6% for functions and 3% for each functionality profile element. We used a CGP grid size of 5 rows by 5 columns, large enough to be expressive yet small enough to discourage over-learning. The function set is defined in Table II.

TABLE II: Function Set

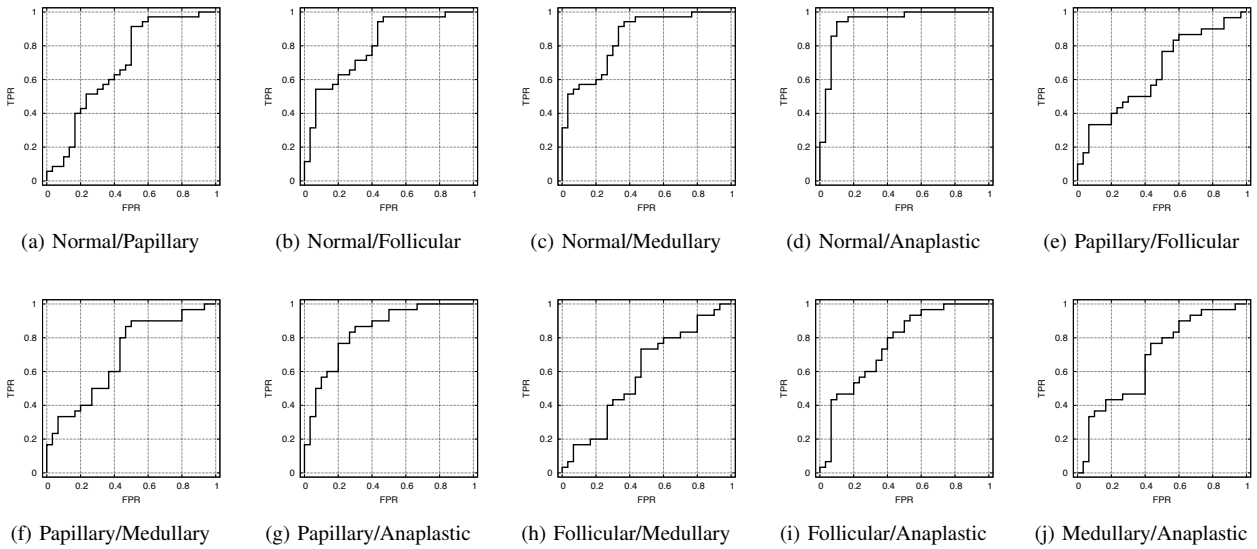
Function	Description
$x + y$	Returns the sum of its two inputs
$x - y$	Returns the difference of its two inputs
$x * y$	Returns the product of its two inputs
$\overline{\{x, y\}}$	Returns the mean of its two inputs
$\min\{x, y\}$	Returns the lesser of its two inputs
$\max\{x, y\}$	Returns the greater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

IV. Results

Figure 3 (left-hand side) shows the distribution of fitness scores for the 50 runs. Whilst the mean multiclass AUC score is 0.68, a number of classifiers were found with scores above 0.7, and analysis of the pairwise AUC scores shows good discrimination ($\geqslant 0.8$) between well-seperated classes. Table III, by way of example, gives the overall and pairwise scores for the highest scoring classifier (corresponding ROC curves are shown in Figure 4): showing good discrimination of follicular, medullary and anaplastic cancers from normal cells, good discrimination of medullary and anaplastic cancers from papillary cancer, and fairly good discrimination between follicular and anaplastic cancers. Low scores are generally associated with neighbouring classes, i.e. normal and papillary, papillary and follicular, follicular and medullary, and medullary and anaplastic.

TABLE III: Multi-class and pairwise AUC scores for classifier with highest test score. AUC scores over 0.75 are shown in bold.

Classes	Train AUC	Validate AUC	Test AUC
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



(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

Fig. 4: ROC Curves, showing the ability of the classifier to discriminate between the different thyroid cell lines.

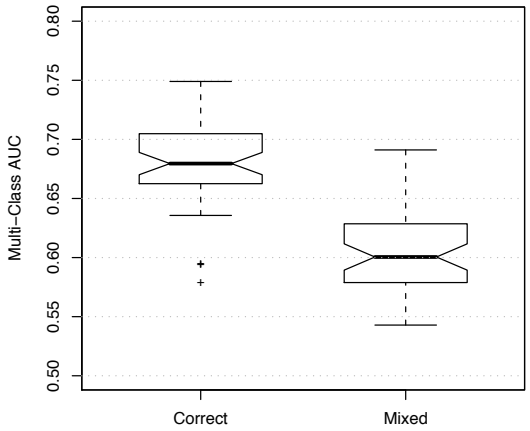


Fig. 3: Comparing the fitness distributions of classifiers evolved to discriminate classes in order of biological aggressiveness (*correct*) with those evolved to discriminate a biologically meaningless ordering (*mixed*).

Although the Raman spectra were normalised with respect to total energy, variability of equipment and experimental conditions could lead to significant local differences in spectral intensity. Whilst the likelihood is considerably mitigated by a multi-class approach, it is possible that an evolved classifier could pick up biochemically-meaningless artefacts whose magnitude happens to be correlated with the ordering of the classes. To test for this, we carried out a further 50 runs with the mixed-up class ordering $< 1, 5, 2, 4, 3 >$, in which biologically-distant classes are neighbouring, i.e. for which we would not expect to be able to evolve meaningful multi-class classifiers. Figure 3 compares the fitness distributions of the two sets of runs, showing that the fitness of classifiers discriminating correctly-ordered classes is, in general, considerably higher than those discriminating the mixed-up ordering. This suggests that high fitness classifiers are making decisions based upon biological factors rather than biologically-meaningless features.

One of the benefits of a GP approach over, for instance, a neural network-based approach, is the relative ease with which solutions can be interpreted. Table IV, for example, lists the expressions used by the 10 highest scoring evolved classifiers. On the whole, these expressions are quite simple. Whilst they vary considerably in form, they often refer to similar regions within the Raman spectrum. This is illustrated by Figure 5, which shows the Raman shifts referenced by these expressions. In particular, it shows that Raman intensities around 636, 833, 1000, 1600 and 1660 are particularly well-referenced. These regions of the spectrum correspond to DNA/RNA nucleotide bases, phenylalanine, aromatic amino acids and amide I, respectively; suggesting that the relative levels of these chemical components may be indicative of thyroid cancers. However, given that Raman spectra capture both qualitative and quantitative information, it is unclear whether these changes are due to the biochemical alteration of native compounds, or due to changes in the concentrations of native chemical components following carcinogenesis. It would be necessary to employ other biochemical analysis methods to determine the actual chemical composition of the cells and thus corroborate the Raman information.

V. Conclusions

We have shown that IRCGP, when applied to Raman spectra, can be used to discriminate between both normal and cancerous thyroid cells, and between different forms of thyroid cancer. We have also shown how analysis of evolved expressions can be used to identify the Raman shifts which underlie good discrimination; providing possible insight into biochemical factors which may be indicative of thyroid cancer.

TABLE IV: Highest scoring evolved classifiers. Numerical values refer to Raman shifts.

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

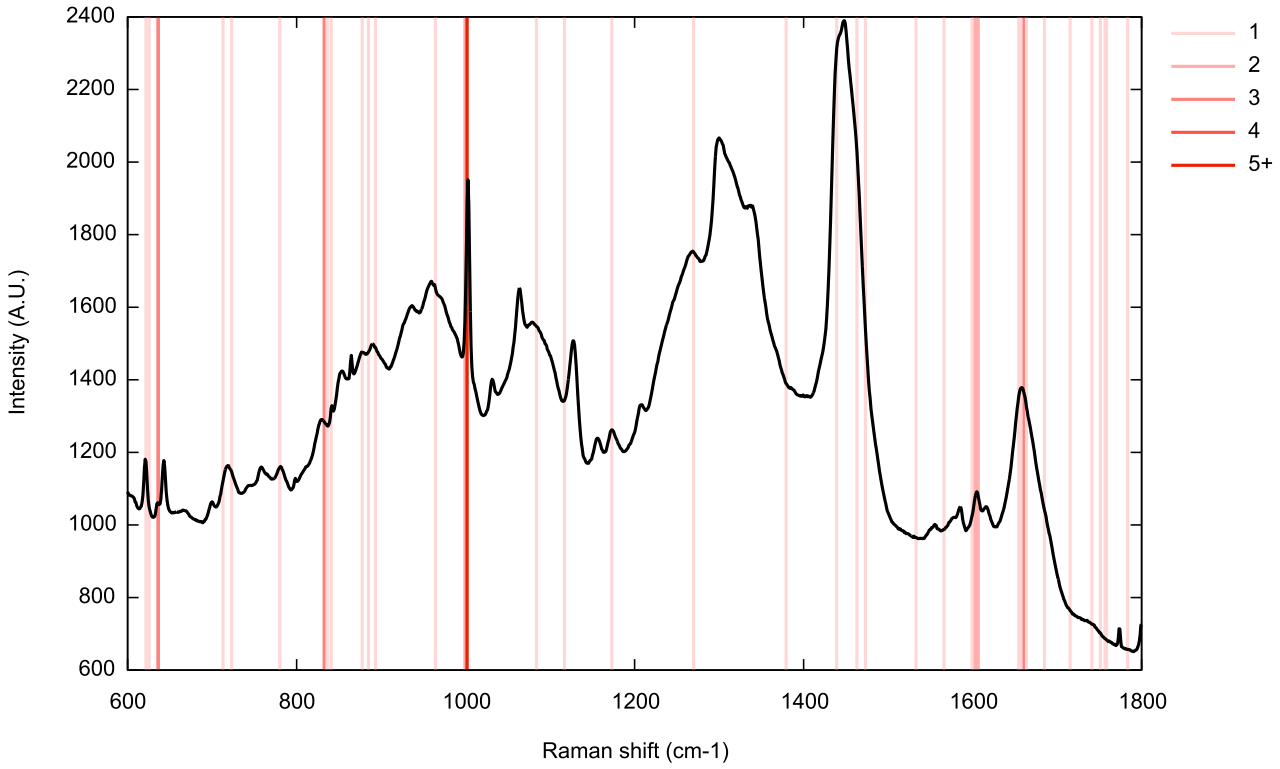


Fig. 5: Mean Raman spectra of all cell samples overlaid with vertical red lines showing Raman shifts referenced by the 10 highest scoring evolved classifiers. Colour intensity is proportional to the number of referring classifiers (see legend).

The stochastic nature of evolutionary algorithms means that multiple runs often lead to multiple, diverse, solutions. In this work, we have leveraged this behaviour in order to identify spectral features which contribute towards good classification. In future work, we plan to extend this approach by looking at whether better classifiers can be produced by combining multiple, diverse, solutions. One way in which this can be done is to use conventional leveraging techniques, such as bagging and boosting, which have been shown able to improve the accuracy and generality of genetic programming [10]. Alternatively, we could co-evolve ensembles alongside classifiers, an approach which proved effective in [16], and which has the advantage of guiding search towards classifiers that work well in ensembles.

REFERENCES

[1] X. Cai, S. L. Smith, and A. M. Tyrrell. Positional independence and recombination in Cartesian genetic programming. In *Proc. 2006 European Conference on Genetic Programming (EuroGP)*, pages 351–360, 2006.

[2] Y. Cao and D. Dasgupta. An immunogenetic approach in chemical spectrum recognition. In A. Ghosh and S. Tsutsui, editors, *Advances in evolutionary computing: theory and applications*, Natural Computing Series, pages 897–914. Springer-Verlag New York, Inc., 2003.

[3] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

[4] R. Goodacre. Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules. *Vibrational Spectroscopy*, 32(1):33–45, 2003.

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

[6] E. B. Hanlon, R. Manoharan, T.-W. Koo, K. E. Shafer, J. T. Motz, M. Fitzmaurice, J. R. Kramer, I. Itzkan, R. R. Dasari, and M. S. Feld. Prospects for in vivo Raman spectroscopy. *Physics in Medicine and Biology*, 45(2):R1, 2000.

[7] A. T. Harris, M. Garg, X. B. Yang, S. E. Fisher, J. Kirkham, D. A. Smith, D. P. Martin-Hirsch, and A. S. High. Raman spectroscopy and advanced mathematical modelling in the discrimination of human thyroid cell lines. *Head & Neck Oncology*, 1:38, October 2009.

[8] A. T. Harris, A. Lungari, C. J. Needham, S. L. Smith, M. A. Lones, S. E. Fisher, X. B. Yang, N. Cooper, J. Kirkham, D. A. Smith, D. P. Martin-Hirsch, and A. S. High. Potential for Raman spectroscopy to provide cancer screening using a peripheral blood sample. *Head & Neck Oncology*, 1:34, September 2009.

[9] K. Hennessy, M. G. Madden, J. Conroy, and A. G. Ryder. An improved genetic programming technique for the classification of Raman spectra. *Knowledge-Based Systems*, 18(4-5):217–224, 2005. ISSN 0950-7051.

[10] H. Iba. Bagging, boosting, and bloating in genetic programming. In W. Banzhaf et al., editors, *Proceedings of the Genetic and Evolutionary Computation Conference*, volume 2, pages 1053–1060. Morgan Kaufmann, 1999.

[11] C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker, and N. Stone. Vibrational spectroscopy: a clinical tool for cancer diagnostics. *Analyst*, 134:1029–1045, 2009.

[12] B. K. Lavine, C. E. Davidson, and A. J. Moores. Genetic algorithms for spectral pattern recognition. *Vibrational Spectroscopy*, 28(1):83–95, 2002.

[13] M. A. Lones. *Enzyme Genetic Programming: Modelling Biological Evolvability in Genetic Programming*. PhD thesis, Department of Electronics, University of York, 2003.

[14] M. A. Lones and A. M. Tyrrell. Biomimetic representation with enzyme genetic programming. *Genetic Programming and Evolvable Machines*, 3(2):193–217, June 2002.

[15] M. A. Lones and A. M. Tyrrell. Modelling biological evolvability: Implicit context and variation filtering in enzyme genetic programming. *BioSystems*, 76(1–3): 229–238, August–October 2004.

[16] M. A. Lones and A. M. Tyrrell. A co-evolutionary framework for regulatory motif discovery. In D. Srinivasan and L. Wang, editors, *2007 IEEE Congress on Evolutionary Computation*, pages 3894–3901. IEEE Press, 2007.

[17] J. F. Miller and P. Thomson. Cartesian genetic programming. In R. Poli et al., editors, *Third European Conf. Genetic Programming*, volume 1802 of *Lecture Notes in Computer Science*, 2000.

[18] S. L. Smith and M. A. Lones. Implicit context representation Cartesian genetic programming for the assessment of visuo-spatial ability. In *Proc. IEEE Congress on Evolutionary Computation 2009*, 2009.

[19] S. L. Smith, S. Leggett, and A. M. Tyrrell. An implicit context representation for evolving image processing filters. In *Proceedings of the 7th Workshop on Evolutionary Computation in Image Analysis and Signal Processing*, volume 3449 of *Lecture Notes in Computer Science*, pages 407–416, 2005.

[20] S. L. Smith, P. Gaughan, D. M. Halliday, Q. Ju, N. M. Aly, and J. R. Playfer. Diagnosis of Parkinson's disease using evolutionary algorithms. *Genetic Programming and Evolvable Machines*, 8(4):433–447, 2007.

[21] C. S. B. Teixeira, R. A. Bitar, H. S. Martinho, A. B. O. Santos, M. A. V. Kulcsar, C. U. M. Friguglietti, R. B. da Costa, E. . L. Arisawa, and A. A. Martin. Thyroid tissue analysis through Raman spectroscopy. *Analyst*, 134:2361–2370, 2009.