

High Dimensionality and Scaling-up Performance of RBF Models with Application to Healthcare Informatics

Julio de Alejandro Montalvo, George Panoutsos, Mahdi Mahfoufand, and James W. Catto

Abstract—In this paper, the scaling-up performance of Radial Basis Function (RBF) Neural-Fuzzy models is investigated. The work presented is based on the challenge of analyzing microarray data for the prediction of the patients' cancer survival. The aim is to find the limit for the maximum number of inputs to use in the model while maintaining low computational complexity and high accuracy. The combination of Fuzzy C-means and RBF-Neural-Fuzzy models presents the challenge of scaling-up when more than a thousand inputs are used. To overcome this challenge we introduce a Weighted Fuzzy C-means (WFCM) algorithm. A second contribution is a cluster optimization algorithm based on the Xie-Beni cluster validity index to improve the quality of the clusters calculated by the WFCM. The best performance (0.87 AUC) for the prediction of the patients' bladder cancer survival was achieved by a 500-gene signature model with a modeling structure having only five rules.

Index Terms—Bladder cancer, feature-selection, fuzzy logic, health-care informatics, high dimensionality low sample size, microarray, neural-fuzzy, radial-basis-function (RBF).

I. INTRODUCTION

The analysis of high dimension-low sample size data (HDLSS) represents a systems engineering classification and identification challenge. This is due to the noisy characteristics of high dimensional data and the fact that the number of replications for the experiment is very low (not enough samples for the model's training algorithm to use). The study presented in this paper is based on the healthcare informatics challenge of analyzing large-scale microarray cancer data (high dimension-low sample size data) for the prediction of the patient's cancer survival outcome. Microarray is a microchip that contains thousands of spots each one filled with genetic information; it is a high throughput technology that is capable of simultaneously evaluating quantitative measurements for the expression of thousands of genes [1]. Generated datasets from each microarray are in the region of tens of thousands of genes per patient sample, thus creating a very complex and high-dimension data-mining task.

Manuscript received August 15, 2014; revised October 29, 2014. This work was supported in part by Conacyt, Mexico.

Julio de Alejandro, George Panoutsos, and Mahdi Mahfouf are with The University of Sheffield, Department of Automatic Control and Systems Engineering, Mappin Street, Sheffield, S1 3JD, UK. Julio de Alejandro and George Panoutsos are also with the Institute for In-Silico Medicine (INSIGNEO), University of Sheffield, UK (e-mail: j.montalvo@sheffield.ac.uk, g.panoutsos@sheffield.ac.uk, m.mahfouf@sheffield.ac.uk).

James W. Catto is with The University of Sheffield, Academic Urology Unit and Institute for Cancer Studies GU22, G Floor, The Medical School, Beech Hill Road, Sheffield, S10 2RX, UK (e-mail: j.catto@sheffield.ac.uk).

To tackle the challenge of high number of features, feature selection algorithms have become indispensable components of the data mining process [2]. There are three categories for feature selection: filters (used as a pre-processing step), wrappers (optimize a predictor as part of the feature selection process) and embedded methods (perform feature selection in the process of training). Generally, filter feature selection methods are used in combination with wrapper methods to diminish the computational cost of examining the complete data set. Statistical regression methods have difficulties to make an accurate prediction when there are multiple interconnected variables and in the presence of contaminating data, such as microarray data [3]. The question raised is if the combination of filter and wrapper methods offers significant advantages in terms of tolerance to imprecision and accuracy in the prediction, compared to using only a wrapper method and a higher number of inputs. The combination of filter-wrapper methods have proven to be an effective method for classification [4].

We can address a number of challenges associated to the theme of this paper; it is important to know if it is possible to avoid the use of feature selection techniques. Specifically, avoid the use of filter-based feature selection techniques that do not assess if there is interdependency in the data, but only assess one-to-one variable dependence. Existing studies suggest that best classification results are obtained by selecting 100-500 genes in a Support Vector Machine (SVM) model [5], [6]. However, is this limitation a result of the modeling characteristics of SVM models or would a different method provide a better outcome?

In this paper, an assessment is performed of the scalability of Radial Basis Function (RBF) Neural-fuzzy models with high dimensionality and low number of samples. We chose an RBF-Neural-Fuzzy system because it offers a good balance of performance and simplicity while being tolerant to some imprecision and crucially being capable of accurate model representations even when few samples are available [7], [8]. The aim is to assess if it is possible to avoid the use of filter-based feature selection methods; and conclude if the proposed modeling approach scales-up (i.e. performs well when the number of genes is increased).

As stated by Koepke [9] "One should not rely on clustering results alone for high dimensional data and one should do feature selection". Clustering is a form of data analysis where the data is divided into groups or subsets where the objects present in that subset share some similarities.

We can divide clustering in two types: hard clustering and fuzzy clustering. Hard clustering refers to an inflexible boundary for the partitions compared to the vagueness showed in fuzzy clustering where a data point can belong to different classes with different membership values [10].

Numerous methodologies have been applied to the problem of clustering HDLSS data, for example: based on p values [11], k-means clustering [12].

In [11] the authors propose a hard clustering algorithm based on p-values as a measure of similarity where no optimization is necessary. Nevertheless, it is believed that fuzzy clustering is a more appropriate method to find clusters due to its robustness to noise [13].

In this paper, the change in the variation of the predictive accuracy of the models, when the number of inputs is increased or reduced, is evaluated using a model for the prediction of survival in bladder cancer [14]. As a pre-input selection we use the t-test statistical method to systematically reduce the large initial dataset. This is a widely applied pre-processing step in microarray gene selection, aiming at eliminating the ‘easy to identify’ and obviously irrelevant to the process genes.

The wrapper method used in the proposed approach, is based on Fuzzy Logic (FL) and a Radial-Basis-Function (RBF) Neural-Fuzzy (NF) computational structure. We chose a hybrid Neural-Fuzzy Logic model for the reason that they have the learning ability of Neural Systems and the interpretability of Fuzzy Systems, and have the ability to automatically generate and adjust the membership functions and linguistic rules directly from the data samples. Three different approaches are used in this study to assess the effectiveness of modeling HDLSS data: a) a Fuzzy C-Means (FCM) and RBF-NF modeling structure; b) a Weighted Fuzzy C-Means (WFCM) and an RBF-NF modeling structure; and c) a Weighted Fuzzy C-Means (WFCM) and an RBF-NF modeling structure with the help of a cluster validation index. All the proposed approaches of the Levenberg-Marquardt [15] algorithm for the model’s parametric optimization.

The remainder of this article is organized as follows: Section II) Methodology: A description of the data-mining and modeling methodology is presented. Section III) Simulation Results: Results are shown for the three different modeling approaches applied to the prediction of survival in a bladder cancer and Section IV) Conclusions.

II. METHODOLOGY

The methodology is organized in three, incremental, parts, whereby a simple FCM-based RBF-NF modeling approach is further enhanced with measures of weighted-clustering followed by a cluster validity approach.

A. FCM and RBF-NF Function Model

The data-mining workflow (Fig. 1) consists of an initial data pre-processing step, where data normalization is performed followed by a student’s distribution t-test to eliminate easy to identify irrelevant to the process genes. The following step consists of applying Fuzzy C-means (FCM) clustering for the creation of the initial rule-base. This rule-base is then ‘translated’ into a Radial-Basis-Function Neural-Fuzzy structure (one multi-dimensional cluster corresponds to one Fuzzy Logic rule), and the modeling structure is finally parametrically optimized via the Levenberg-Marquardt function-minimization algorithm [15].

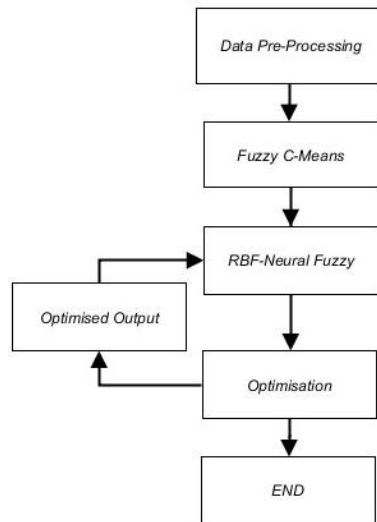


Fig. 1. Data-mining workflow for the FCM and RBF-NF model.

The FCM method [16] is frequently used in pattern recognition but the main reason to use it here is because the results of Fuzzy C-Means (membership degree per sample) can be used directly as initial values (rule base) of an RBF Model. The following FCM minimization objective function is used:

$$J_{m=1}^M = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, 1 \leq m < \infty \quad (1)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d -dimensional measured data, c_j is the d -dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center.

The next stage consists of introducing an RBF function to describe a Neural-Fuzzy system [17]. Fig. 2 shows the structure of the RBF-NF model, where the input, rule-base (hidden layer) and output layers can be identified. The presented system can then be parametrically optimized via a suitable function minimization algorithm. The RBF-NF modeling structure can inherently represent knowledge in a linguistic format that can be easily interpreted by clinicians.

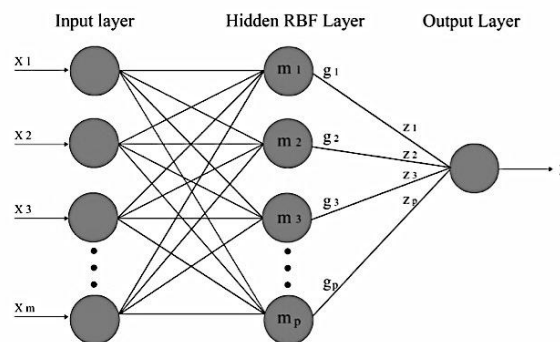


Fig. 2. RBF Layers of NF model.

This is in the form of simple linguistic rules-sentences (IF-THEN rules).

The linguistic statements are in the form:

If Gene 1 has x intensity and Gene 2 has y intensity **Then** the Patient’s Outcome is Z .

The intensities of the genes are equally divided in 7

categories according to their value: Very Low, Low, Low Medium, Medium, Medium High, High and Very High.

In our proposed framework, the consequent part of the Fuzzy logic system, (...Then the Patient's Outcome is...) is defined as a linear function (Takagi-Sugeno-Kang, TSK), which is a linear combination of the model's inputs.

Finally, the model is parametrically optimized via the Levenberg-Marquardt (LM) algorithm. The LM algorithm is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions [15].

B. WFCM and RBF-NF Function Model

FCM algorithms consider each object equally important in the cluster solution. For that reason, when FCM is applied to a high number of inputs (more than a thousand), the rule-base loses clarity due to the high dimensional space and the values of the membership degree become truly small. The problem that arises is that the FCM clusters are the initial conditions for the RBF Neural-Fuzzy and because of their poor quality, the optimization algorithm fails. By applying Weighted FCM (WFCM) we define the relative importance of each object to the clustering solution. We applied this weighted factor to the output of the data to improve the membership degree of each cluster. This modification improves the initial condition of the RBF Neural-Fuzzy. The second enhancement presented in this paper (Fig. 3) consists of applying a Weighted Fuzzy C-means clustering for the creation of the initial rule-base and applying the rule-base directly to the RBF Neural-Fuzzy. The rule-base is then 'translated' into a Radial-Basis-Function Neural-Fuzzy structure, and is parametrically optimized via the Levenberg-Marquardt function minimization algorithm.

The weighted FCM (WFCM) is based on the minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m w_i \|x_i - c_j\|^2, 1 \leq m < \infty \quad (2)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j , x_i is the i th of d -dimensional measured data, c_j is the d -dimension center of the cluster, $\|\cdot\|$ is any norm stating the likeness between the measured data and the center and w_i is a weighted factor applied to the output of the data and is equal to the number of inputs.

The membership u_{ij} and the cluster centers c_j are calculated by:

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}, \quad (3)$$

$$c_j = \frac{\sum_{i=1}^N w_i u_{ij}^m * x^j}{\sum_{i=1}^N w_i u_{ij}^m}$$

Each sample will have a membership assigned (u_{ij}) in every cluster; a higher membership would translate into a higher degree of similarity between the sample and the cluster. Each derived information granule (data-cluster) depicts a process rule in the Fuzzy Logic domain. The weighted FCM is similar to the one proposed in [18], [19], however in this paper the weighting factor changes according

to the number of inputs that are used for the model. The purpose is to analyze the scaling-up predictive performance of an RBF-NF model applied to the prediction of bladder cancer's survival.

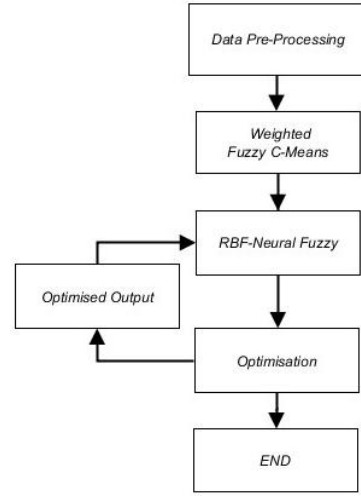


Fig. 3. Data-mining workflow for the WFCM and RBF-NF model.

WFCM, validation index and RBF-NF function model as the dimensionality of the RBF structure increases the data-mining of the initial rules and separately between the MF's becomes problematic. In this section we introduce a cluster-validity index to the data-mining process to further improve the quality of the rule-base. Fig. 4 depicts the validity index data-mining workflow. There are multiple indices for validation of the fuzzy clusters; partition coefficient [20], partition entropy [21], Fukuyama and Sugeno [22], Xie-Beni [23]. Most of the validation indices aim to find the optimal number of clusters, but in this paper we employ this index to improve the quality of the clusters calculated by the WFCM. We use a modification of the Xie-Beni index, as presented in [18]. A reliable validation index should take into consideration the compactness or how close each point of the cluster is and the separation of the FMC clusters, which is the case in the Xie-Beniindex (4).

$$Id = \frac{\sum_{k=1}^N \sum_{j=1}^C w_k (u_{kj}) \|v_k - v_j\|^2}{n \min_{j \neq i} \{ \|v_j - v_i\|^2 \}} \quad (4)$$

The measure of Compactness (C) is given by:

$$C = \frac{\sum_{k=1}^N \sum_{j=1}^C w_k (u_{kj})^m \|v_k - v_j\|^2}{n} \quad (5)$$

The measure of separation is given by:

$$Separation = \min_{j \neq i} \{ \|v_j - v_i\|^2 \} \quad (6)$$

where C is the number of clusters, u_{kj} is the membership degree, w_k is the weight of significance assigned to v_k , which is the complete data, and v_i are the centers of the clusters. The optimal partition clusters would have to be as compact as possible, while they maintain a good balance between separation and coverage of the input space [10]; these characteristics would translate into a high quality

rule-base.

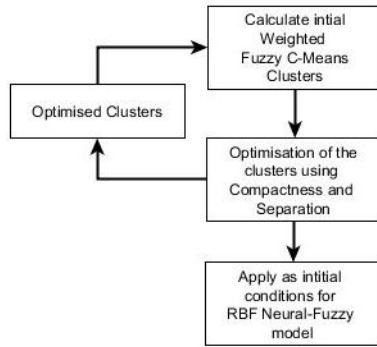


Fig. 4. Flow chart of the processing of the data with weighted FCM and the validation index.

III. SIMULATION RESULTS

In this section the healthcare-based case study is first introduced, and then the scaling-up simulation results of the modeling methodology are presented.

A. Microarray Model-Based Prediction of Patients' Cancer Outcome

The introduction of microarray-based technologies for analyzing patient tissues has produced a significant challenge for healthcare clinicians as well as data analysts; the challenge of understanding and using efficiently thousands of gene-based data and linking them to clinically useful information. The case study presented in this paper is focused on bladder cancer microarray data, specifically the ones presented in the Sanchez-Carbayo study [14]. In Table I a summary of the data is presented.

TABLE I: BLADDER CANCER – MICROARRAY GENE INTENSITY DATA SETS

Data Set	Microarray platform	Number of genes	Number of Samples (patients)
Sanchez-Carbayo	Affymetrix U133A	22283	90

For modeling purposes the survival outcome of the patients was numerically encoded as '-1' for 'No Evidence of Disease – NED' and '1' for 'Dead of Disease - DOD'.

The data samples were randomly separated into 'training' (70% of the patients) and 'testing' (30% of the patients) datasets. The training set is only used to train the model, and the testing dataset is only used after the model training is finished to test the generalization performance of the model, as a form of cross-validation [24].

In modeling such a dataset, gradually increasing the number of inputs (genes used in the model) also would increase the computational requirements of the process – this may or may not be an issue depending on the application. However, does a larger more complex model (in terms of number of inputs and structure) correspond to an enhanced performance? In the following section (scaling-up performance of RBF-NF models) we present a comparison between models of 100 to 5000 genes. During this comparison, we identify a number of computational and model performance-related challenges, and we show how the introduction of the proposed data mining and modeling framework helps resolve such challenges. The training time of each model depends on the number of samples and inputs.

On average, using a standard PC with an Intel ® Core™ i7 CPU 870 @ 2.93 GHz processor with 8GB of RAM, it takes a few minutes to process (train, validate) 100 inputs. The computational requirements increase dramatically, as the number of genes is also increased, to more than 24 hours for 1000 genes. For the models that use 2000 to 5000 genes it was necessary to make use of a High Performance Computing (HPC) server with multiple computing cores [25].

B. Scaling-up Performance of RBF-NF Models

The RBF-NF model was developed as described in section II. The methodology was applied to the data set of Sanchez-Carbayo [14] for the prediction of survival in bladder cancer. The question raised at the beginning of this paper is if the combination of filter and wrapper methods offers significant advantages in terms of tolerance to imprecision and accuracy in the predictions, compared to using only a wrapper method and a higher number of inputs.

The results shown in this section confirm if it is possible to avoid the use of feature selection techniques. In the case that this premise is true, the rational limit for the maximum number of inputs to use in the model needs to be established. In terms of the modeling structure, we maintain five (5) fuzzy rules throughout the modeling study for comparison purposes. Based on previous research work[8] five rules in this case study offers a good balance of performance and model simplicity.

1) Results with 100 inputs and 5 rules

The methodology was applied to the data set of Sanchez-Carbayo for the prediction of survival of in bladder cancer. The results shown in Table II are the Area Under the curve (AUC) of the models (ROC analysis). The FCM has the highest performance compared to the other two WFCM models. The difference in performance between the two models using WFCM is however not significant.

TABLE II: PERFORMANCE OF THE MODEL USING 100 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC	AUC	AUC	AUC	AUC	AUC
100	Train	Test	Train	Test	Train	Test
	1.00	0.86	1.00	0.80	1.00	0.79

2) Results with 300 inputs and 5 rules

TABLE III: PERFORMANCE OF THE MODEL USING 300 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC	AUC	AUC	AUC	AUC	AUC
300	Train	Test	Train	Test	Train	Test
	1.00	0.87	1.00	0.81	1.00	0.82

The results shown in Table III are the Area Under the curve (AUC) of the model with 300 inputs. In the same manner that the models behave with 100 inputs, the highest performance was obtained by the model using FCM and the difference in performance between the two WFCM models is

not significant. Moreover, we can perceive a trend for all the models of increase for the AUC.

3) Results with 500 inputs and 5 rules

The results shown in Table IV include the Area Under the curve (AUC) of the model with 500 inputs.

TABLE IV: PERFORMANCE OF THE MODEL USING 500 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC		AUC		AUC	
	Train	Test	Train	Test	Train	Test
500	1.00	0.73	1.00	0.75	1.00	0.87

The WFCM model with the validation index for the initial clusters outperforms the WFCM model that did not use the validation index and the FCM model. The WFCM model with the validation index keeps the same trend of an increase in the AUC while the AUC for the WFCM and the FCM model start having a decrease in the performance.

4) Results with 1000 inputs and 5 rules

TABLE V: PERFORMANCE OF THE MODEL USING 1000 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC		AUC		AUC	
	Train	Test	Train	Test	Train	Test
1000	0.96	0.65	1.00	0.69	1.00	0.80

Similar to the results obtained for the model with 500 inputs, the WFCM model with the validation index for the initial clusters clearly outperforms the WFCM model and the FCM model (Table V). The WFCM model with the validation index now had a decrease in the AUC, the same case presents for the WFCM and the FCM.

5) Results with 2000 inputs and 5 rules

As discussed earlier in the paper, the FCM fails to converge as the complexity increase to more than 2000 genes. This is noted as 'N/A' in the Table VI.

TABLE VI: PERFORMANCE OF THE MODEL USING 2000 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC		AUC		AUC	
	Train	Test	Train	Test	Train	Test
2000	N/A	N/A	0.75	0.67	0.75	0.73

The WFCM models have similar performance for training however testing performance is higher for the WFCM model with the validation index. The results also show a general trend of decrease in the performance for both models.

6) Results with 5000 inputs and 5 rules

Both models have a significant decrease in the AUC, however they perform the same. Overall, up to around 300 genes, a simple FCM clustering technique is adequate to resolve the modeling complexity of RBF modeling structures.

As the number of genes increases, but number of samples remains the same, the WFCM and WFCM with the validity index are needed to model the gene microarray data with a good level of performance. Above 500 genes the WFCM with the validity index starts to outperform the WFCM; however the modeling structure appears to reach its limit in terms of resolving complexity above 5000 genes, where there is a dramatic drop in performance.

TABLE VII: PERFORMANCE OF THE MODEL USING 5000 INPUTS AND 5 RULES

Genes	FCM		WFCM		WFCM and validation index	
	AUC		AUC		AUC	
	Train	Test	Train	Test	Train	Test
5000	N/A	N/A	0.57	0.52	0.57	0.52

IV. CONCLUSION

In this paper the scaling-up performance of Radial Basis Function (RBF) Neural-Fuzzy models is investigated. RBF-Neural-Fuzzy models offer balance of performance and simplicity (while being tolerant to imprecision); these are traits that are important in healthcare informatics. An enhanced rule-base extraction framework was proposed to improve the model's performance for high-dimensional low sample size data. The work presented in this paper is based on the healthcare informatics challenge of analyzing large-scale microarray cancer data for the prediction of the patients' cancer survival outcome. The simulations obtained for the prediction of bladder cancer's survival provides a better understanding of the scalability performance for RBF Neural-Fuzzy models.

From the results obtained we can conclude that the RBF model using FCM alone performs best when less than 300 genes are used. Due to the characteristics of high-dimension low sample size data, as the number of genes increases but number of samples remains the same, the WFCM and WFCM with the validity index are needed to model the microarray data with a good level of performance. Above 500 genes the WFCM with the validity index starts to outperform the WFCM. A dramatic drop in the performance is observed above 5000 genes, where the modeling structure appears to reach its limit in terms of resolving complexity. Maximum accuracy for the prediction was obtained by using five hundred inputs for the WFCM and the validation index (0.87 AUC) and three hundred inputs for the FCM (0.87 AUC).

The developed models maintain the simple structure with just five (5) rules, but with very good performance (up to 2000 genes). The simple linguistic-based structure of the Fuzzy-logic system could be used in human-centric decision support systems. One need to keep in mind that the training time for the models can still be up to 3-4 days on a high performance computing server, however other –more efficient- optimization algorithm can be used instead.

REFERENCES

[1] E. Dudzic *et al.*, "Integrated epigenome profiling of repressive histone modifications, DNA methylation and gene expression in normal and malignant urothelial cells," *PLoS One*, vol. 7, p. e32750, 2012.

- [2] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: A study on high-dimensional spaces," *Knowledge and Information Systems*, vol. 12, pp. 95-116, 2007.
- [3] H. B. Burke *et al.*, "Artificial neural networks improve the accuracy of cancer survival prediction," *Cancer*, vol. 79, pp. 857-862, 1997.
- [4] N. R. Pal *et al.*, "Discovering biomarkers from gene expression data for predicting cancer subgroups using neural networks and relational fuzzy clustering," *BMC Bioinformatics*, vol. 8, p. 5, 2007.
- [5] T. Li, C. Zhang, and M. Ogihara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," *Bioinformatics*, vol. 20, pp. 2429-2437, 2004.
- [6] C. D. H. Chai, "An evaluation of gene selection methods for multi-class microarray data classification," in *Proc. the workshop W9 on data mining and text mining for bioinformatics*, 2004, pp. 3-10.
- [7] S. M. Samuri *et al.*, "Neural-fuzzy modelling of lung volume using absolute electrical impedance tomography," in *Proc. International Conference on Bio-Inspired Systems and Signal Processing, BIOSIGNALS*, Rome, 2011, pp. 43-50.
- [8] J. de A. Montalvo, *et al.*, "Radial basis function neural-fuzzy model for microarray signature identification," in *Proc. the International Conference on Bioinformatics Models, Methods and Algorithms*, 2013, pp. 134-139.
- [9] H. A. Koepke and B. S. Clarke, "On the limits of clustering in high dimensions via cost functions," *Stat. Anal. Data Min.*, vol. 4, pp. 30-53, 2011.
- [10] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognition Letters*, vol. 19, pp. 237-246, 1998.
- [11] G. V. Borries and H. Wang, "Partition clustering of high dimensional low sample size data based on p-values," *Comput. Stat. Data Anal.*, vol. 53, pp. 3987-3998, 2009.
- [12] S. Tavazoie *et al.*, "Systematic determination of genetic network architecture," *Nat Genet.*, vol. 22, pp. 281-285, 1999.
- [13] T. Hanai, H. Hamada, and M. Okamoto, "Application of bioinformatics for DNA microarray data to bioscience, bioengineering and medical fields," *J. Biosci Bioeng.*, vol. 101, pp. 377-384, 2006.
- [14] M. Sanchez-Carbayo *et al.*, "Defining molecular profiles of poor outcome in patients with invasive bladder cancer using oligonucleotide microarrays," *Journal of Clinical Oncology*, vol. 24, pp. 778-789, 2006.
- [15] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *The Quarterly of Applied Mathematics*, vol. 2, pp. 164-168, 1944.
- [16] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *Cybernetics and Systems*, vol. 3, pp. 32-57, 1973.
- [17] G. Panoutsos and M. Mahfouf, "A neural-fuzzy modelling framework based on granular computing: Concepts and applications," *Fuzzy Sets and Systems*, vol. 161, pp. 2808-2830, 2010.
- [18] A. D. Niros and G. E. Tsekouras, "On training radial basis function neural networks using optimal fuzzy clustering," in *Proc. 17th Mediterranean Conference on Control and Automation*, 2009, pp. 395-400.
- [19] T. C. Havens *et al.*, "Fuzzy c-Means algorithms for very large data," *IEEE Transactions on Fuzzy Systems*, vol. 20, pp. 1130-1146, 2012.
- [20] J. C. Bezdek, "Numerical taxonomy with fuzzy sets," *Journal of Mathematical Biology*, vol. 1, pp. 57-71, 1974.
- [21] J. C. Bezdek, "Cluster Validity with Fuzzy Sets," *Journal of Cybernetics*, vol. 3, pp. 58-73, 1973.
- [22] Y. Fukuyama and M. Sugeno, "A new method of choosing the number of clusters for the fuzzy c-means method," in *Proc. 5th Fuzzy Syst. Symp.*, 1989.
- [23] X. Xuanli Lisa and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 841-847, 1991.
- [24] U. M. Braga-Neto and E. R. Dougherty, "Is cross-validation valid for small-sample microarray classification?" *Bioinformatics*, vol. 20, pp. 374-380, 2004.
- [25] The University of Sheffield. (May 2014). *Sheffield WRGRID-ICEBERG*. [Online]. Available: <https://www.shef.ac.uk/wrgrid/iceberg>



Julio de Alejandro Montalvo was born in Mexico, on 26 October 1984. He got B.S. degree in electronics and automation, in Universidad Autonoma de Nuevo Leon, San Nicolas de los Garza, Nuevo Leon, Mexico, 2006. In 2010, Julio started his PhD in the Department of Automatic Control and Systems Engineering in The University of Sheffield, United Kingdom. Currently he is working towards the conclusion of his thesis. He worked for The Trane Company from 2007 to 2010 in the Department of Building Automation Systems as a project engineer. His current research interests are high dimensional low sample size data, feature-selection and computational intelligence.



George Panoutsos received a BEng. (Hons) degree in electrical and electronic engineering with microelectronics from the University of Northumbria, Newcastle, UK. He obtained MSc as well as Ph.D. degrees in automatic control and systems engineering from the University of Sheffield, Sheffield, UK. Since November 2009 he has been a lecturer in computational intelligence in the Department of Automatic Control and Systems Engineering, The University of Sheffield. His research interests include computational intelligence, data-driven modeling, decision support systems, data fusion and process optimization in the field of biomedical engineering and healthcare as well as intelligent manufacturing systems.



Mahdi Mahfouf received the M.Phil. and Ph.D. degrees from the University of Sheffield, Sheffield, U.K., in 1987 and 1991, respectively. During 1992-1997, he was a postdoctoral research fellow in the Department of Automatic Control and Systems Engineering, University of Sheffield, where he rejoined the Department of Automatic Control and Systems Engineering as a lecturer in systems and control engineering in October 1997, and was promoted to a senior lecturer in 2000, a reader in 2003, and a personal chair in intelligent systems engineering in 2005. He was involved in the areas of intelligent control and modeling applied to the life sciences and the process industries for more than 20 years. He is the author or co-author of more than 200 papers published in various international journals. His research interests include predictive control, fuzzy/neuro fuzzy systems, intelligent-systems-based modeling and control, evolutionary computing, modeling and control for biomedicine, and modeling and optimization in the metal processing industry. Dr. Mahfouf was the recipient of the prestigious Institution of Electrical Engineers (IEE) HARTREE Premium Award for a paper published in the IEE Proceedings on Control Theory and Applications.



James Catto is a professor in Urological Surgery at the University of Sheffield and an honorary senior clinical research fellow at the University of Oxford. He qualified from Leeds University in 1994 and trained in York and Sheffield, UK before a fellowship in Brisbane, Australia. He was appointed as an honorary consultant Urological Surgeon in 2007 and specializes in Uro-Oncology. He obtained a PhD in 2004 and now runs a research group investigating the biology of urological cancers. He was awarded the European Association of Urology thesis prize and the Bard medal from BAUS for his PhD studies. James currently runs a lab in the Academic Unit of Molecular Oncology. They research the translational application of molecular biology to urological malignancies, and in particular bladder and prostate cancer. James's particular interest is in the epigenetic alterations seen within these tumors.