Evolving Classification Models for Prediction of Patient Recruitment in Multicentre Clinical Trials Using Grammatical Evolution

Gilyana Borlikova[1(✉)], Michael Phillips[2], Louis Smith[2], and Michael O'Neill[1]

[1] Natural Computing Research and Applications Group, School of Business, University College Dublin, Dublin, Ireland
{gilyana.borlikova,m.oneill}@ucd.ie
[2] ICON Plc, Dublin, Ireland
{michael.phillips,louis.smith}@iconplc.com

Abstract. Successful and timely completion of prospective clinical trials depends on patient recruitment as patients are critical to delivery of the prospective trial data. There exists a pressing need to develop better tools/techniques to optimise patient recruitment in multicentre clinical trials. In this study Grammatical Evolution (GE) is used to evolve classification models to predict future patient enrolment performance of investigators/site to be selected for the conduct of the trial. Prediction accuracy of the evolved models is compared with results of a range of machine learning algorithms widely used for classification. The results suggest that GE is able to successfully induce classification models and analysis of these models can help in our understanding of the factors providing advanced indication of a trial sites' future performance.

Keywords: Clinical trials · Enrolment · Grammatical evolution · Grammar-based genetic programming

1 Introduction

Patient recruitment is a major bottleneck in conducting clinical trials [1,2]. As Chris Trizna writes in Chap. 19 of "Re-engineering clinical trials" [1] "No patients - No Data" [3]. The recent Tufts Center for the Study of Drug Development report [4] on patient enrolment shows, that though 89 % of trials eventually enrol the required number of patients, the timelines are usually pushed to nearly twice the original plan. Forty eight percent of sites in a given trial fail to enrol required number of patients resulting in the need to bring more sites into the study and extending overall enrolment timelines. Failure to enrol required number of subjects is one of the most frequent reasons for trials' discontinuation [5]. This situation makes optimisation of patient recruitment "a million dollar question" for pharma and CRO industries. In the work presented here, Grammatical Evolution [6,7], a grammar-based Genetic Programming system [8], was used to induce classification models for prediction of patient enrolment performance

G. Squillero and P. Burelli (Eds.): EvoApplications 2016, Part I, LNCS 9597, pp. 46–57, 2016.
DOI: 10.1007/978-3-319-31204-0_4

of investigators/sites in clinical trials. Though results of the best GE model selected based on the training fitness are below the results of the comparator models, the overall best test results obtained by GE models are comparable or even better than results of the comparator models. However, further work is needed to establish an approach to ensure a better way to select the most generalisable model out of the list produced by independent GE runs. The evolved GE models use only a subset of the predictor variables for classification and are human-interpretable, thus providing an insight into the factors behind the classification. These results illustrate applicability of GE and, more broadly, of evolutionary computation to real world business analytics problems. The rest of the paper is organised as follows. Section 2 provides a brief background to the problem and an overview of the related literature. Section 3 describes the dataset and design of the experiment. Section 4 describes and analyses the results, and Sect. 5 draws conclusions and future work directions.

2 Background

2.1 Subject Enrolment in Clinical Trials

Several components contribute to the success or failure of patient enrolment, such as investigator/site selection, complexity of the trial protocol and different strategies of patient engagement [1,2]. While a site/investigator may be good at recruiting one patient population, they may be not so good at another patient population. A non-performing site in one study may not be a non-performing site for all indications or therapeutic areas. Recently much attention has been given to the better ways of reaching target patient populations and maintaining patient engagement [1]. There is also a growing recognition that excessive complexity of trial protocols adversely affects subject enrolment and retention and attempts were made to better manage trial protocol complexity [1]. However, the process of improving patient enrolment remains an area of active business interest. Most trials are set up through a company's site selection function. Though a lot of empirical knowledge is usually accumulated by the professional site selection analysts, a lot of decisions are still made based on an individual analyst's experience and the use of online trial intelligence resources, rather than advanced analytics tools. At the same time the healthcare industry is gradually starting to adapt predictive business analytics techniques to improve processes and boost performance. There exists an urgent business need to capitalise on the advances in machine learning to develop methods able to address challenges of patient enrolment.

2.2 Different Approaches to Patient Enrolment Prediction

Most published research into patient enrolment concerns modelling enrolment rates and forecasting times that will take achieving certain number of enrolled patients. [9] developed detailed Gamma-Poisson mixture models of patient enrolment in multi-centre studies. There is also a considerable amount of research

related to identifying patients and predicting potential enrolment eligibility from analysis of the existing patient databases [10] and electronic healthcare records [11]. A substantial number of various clinical trial recruitment support systems (CTRSS) was developed over the years utilising different technologies and algorithms and [12] provide a recent review of these systems. However, as [13] conclude after a systematic review of models to predict recruitment to multicentre clinical trials development of new better models is required.

2.3 Grammatical Evolution for Classification

Classification is one of the most used methods in machine learning and data mining [14]. The classification method consists in predicting the value of a categorical attribute (the class) based on the values of other attributes (predicting variables). An evolutionary learning technique of Genetic Programming (GP) [15,16] has been successfully applied to a variety of classification problems [17]. Grammatical Evolution (GE) [6,7] is a grammatical approach to GP [8]. In addition to the many features of GP that make it a very convenient technique for classification, use of grammar in GE allows for extra control of the syntax of evolved programs [18]. Previously GE was successfully applied to evolve classifiers for bond ratings from raw financial information, predict corporate failure and credit risk classification [19–21]. The evolved classifiers were competitive with the results produced by Neural Nets and the GE methodology was suggested to have general utility for rule-induction applications. This study extends GE methodology into the domain of prediction of patient enrolment in multicentre clinical trials.

2.4 Scope of Research

The goal of this study is to produce predictive models of future patient enrolment performance of investigators/sites to be selected to participate in a multicentre clinical trial. More specifically, we are interested in employing GE to evolve binary classification model capable of predicting future performance of investigators/sites. In this first study we use the raw unprocessed dataset with only a minor data preparation in order to establish the baseline of the possible model performance and leave a more sophisticated data pre-processing for the future experiments. We also conduct all experiments using only one cut of the data into the training and test set and will assess robustness of the produced models using multiple training/test splits in the future work.

3 Experimental Design

3.1 Model Data

The dataset used in all experiments was constructed based on a subset of the de-identified historical operational data provided by ICON Plc. on 21 Diabetes

Mellitus Type II Phase III trials. At the first stage of data preparation records with missing values were removed, as well as a few predictor variables with near-zero variance. The resultant dataset consisted of 1233 investigator/site related records and 42 predictor variables. The dataset contained 35 numerical variables and 7 categorical variables describing different aspects of investigator/site. The reference label divided all investigators/sites into two classes based on their patient enrolment performance. Prior to the beginning of experiments the balanced split of the data into a training and testing subsets (60/40 %) was performed using `createDataPartition` function of the CARET package in R [22]. In all experiments model training and tuning was performed using the training subset and then performance of the best models was tested on the test data subset to ascertain how well the evolved models generalise to unseen data.

3.2 Evolutionary Model Representation and Run Parameters

We decided to adopt a decision-tree type approach to constructing the GE classification model. The GE grammar used the function and terminal set detailed in Table 1. For this experiment we intentionally confined the function set to arithmetic operations (including protected division) to cover only linear transformations of the variables. Figure 1 shows the grammar used in the experiment.

Table 1. Function and Terminal sets of GE classifier

Function set	Terminal set
+, -, *, /, and, or, nor, xor, nand	35 numerical predictive variables: x1, ..., x35
equals, not_equals	3 categorical predictive variables: x36, x37, x38
less, greater, less_e, greater_e	4 Boolean predictive variables: x39, ..., x42
	21 random constants in -1.0, ..., 1.0 with 0.1 step

Table 2 details the evolutionary parameters used. The population was initialised using Sensible Initialisation [23] and the maximum derivation tree depth was set to 12. Invalid individuals were handled by reselection. For the Random Search (RS) experiment GE was run with 100 % crossover and 100 % mutation and elitisim (as below). Throughout the experiments classification error (number of misclassified records) was used to measure fitness.

3.3 Performance Measurement and Benchmarking

As the first step, results of the GE model were compared with the results returned by RS run under similar settings. Best and average fitness within each generation were used as a performance measure during the evolutionary run. Performance of the best models evolved over 50 generations was assessed by classification of the previously unseen test set. To benchmark the best evolved GE classification model its performance was compared to performance of a number of

```
<s> ::= if <pred> <out> <out>
<out> ::= <s> | <class>
<class> ::= 0 | 1

<pred> ::=  <bool_bool_comp> <pred> <pred>
         | <bool_num_comp> <expr_num> <expr_num>
         | <bool_bool_comp> <expr_bool> <expr_bool>

<bool_bool_comp> ::= and | or | nor | xor | nand

#Numerical
<bool_num_comp> ::= less | greater | less_e | greater_e
<expr_num> ::= <op_num> <expr_num> <expr_num> | <op_num> <var_num> <var_num> | <var_num>
<op_num> ::= + | - | * | /
<var_num> ::= <const_num> | <ft_num>
<ft_num> ::= x1|x2|x3|x4|x5|x6|x7|x8|x9|x10|x11|x12|x13|x14|x15|x16|x17|x18|x19|x20
            |x21|x22|x23|x24|x25|x26|x27|x28|x29|x30|x31|x32|x33|x34|x35
<const_num> ::= -1|-0.9|-0.8|-0.7|-0.6|-0.5|-0.4|-0.3|-0.2|-0.1
               |0.0|0.1|0.2|0.3|0.4 |0.5|0.6|0.7|0.8|0.9|1

#Categorical
<op_cat> ::= equals | not_equals
<var_cat> ::= <x36> | <x37> | <x38>
<x36> ::= 1|2|3|4|5|6|7|8|9|10|11|12|13|14|15|16|17|18|19|20|21
         |22|23|23|24|24|25|26|27|28|29|30|31|32|33|34|35|36|37
<x37> ::= 1|2|3
<x38> ::= 1|2|3|4|5

#Boolean
<expr_bool> ::= <op_bool> <expr_bool> <expr_bool>
              | <op_bool> <ft_bool> <ft_bool>
              | <ft_bool>
              | <op_cat> <var_cat>
<op_bool> ::= and | or | nor | xor | nand
<ft_bool> ::= x39 | x40 | x41 | x42
```

Fig. 1. Grammar used to construct a GE classifier

Table 2. Evolutionary parameter settings

Parameter	Value
Initialisation	sensible initialisation
Number of runs	30
Population size	1000
Number of generations	50
Selection	tournament
Tournament Size	5 (0.5 % of population size)
Replacement	generational
Elite size	1
Crossover	single point
Crossover Probability	0.9
Mutation	integer flip
Mutation Probability	1 event per individual
Max derivation tree depth	12

well-established machine learning algorithms often used in classification problems. The R CARET package [22] was used to train and tune the models and then to test their performance on the test set. Table 3 contains details of model settings. For illustration, performance of the best variant of each model during training is presented in Fig. 4A. The performance on the previously unseen test set (Fig. 4B and Table 4) was used to compare models using several statistics in addition to classification accuracy.

Table 3. Benchmark machine learning model settings

Model	CARET method	Parameter setting
Support vector machines with radial basis function kernel	svmRadial	sigma=0.01231675, cost=0.25
Classification and Regression Tree (CART)	rpart	Complexity parmeter=0
Multivariate Adaptive Regression Splines (MARS)	gcvEarth	product degree=1
Random Forest (RF)	rf	#randomly selected predictors=7
Nearest shrunken centroids (NSC)	pam	shrinkage threshold=2.231067
Simple Classification Rules (OneRule)	JRip (as per RWEKA)	-

4 Results and Analysis

We performed 30 independent evolutionary runs to evolve GE classification models. Results of this experiment are presented in Fig. 2. The best and average population fitness gradually improved over 50 GE generations. Figure 2B shows results of RS run for the same number of times. Best fitness of GE classifiers start to outperform results of random search from generation 8. This is further confirmed by testing evolved best individuals from generation 50 on the previously unseen test data (Fig. 2D). Though there is a substantial overlap in the test results between RS and GE individuals, the median fitness of GE individuals is substantially better than median fitness of individuals produced by RS. GE was also able to evolve the model which achieves the best classification overall.

When classification quality of the best individuals returned by 30 runs was assessed on the previously unseen test set it became apparent that different individuals returned best results on these two data subsets. The best individual as assessed by training fitness (0.273) performed rather poorly on the test set (test fitness 0.371); while another individual with a lesser training fitness (0.289) demonstrated the best test classification performance (test fitness 0.312). These results suggest a possibility of overfitting [21,24]. Figure 3 presents phenotypes

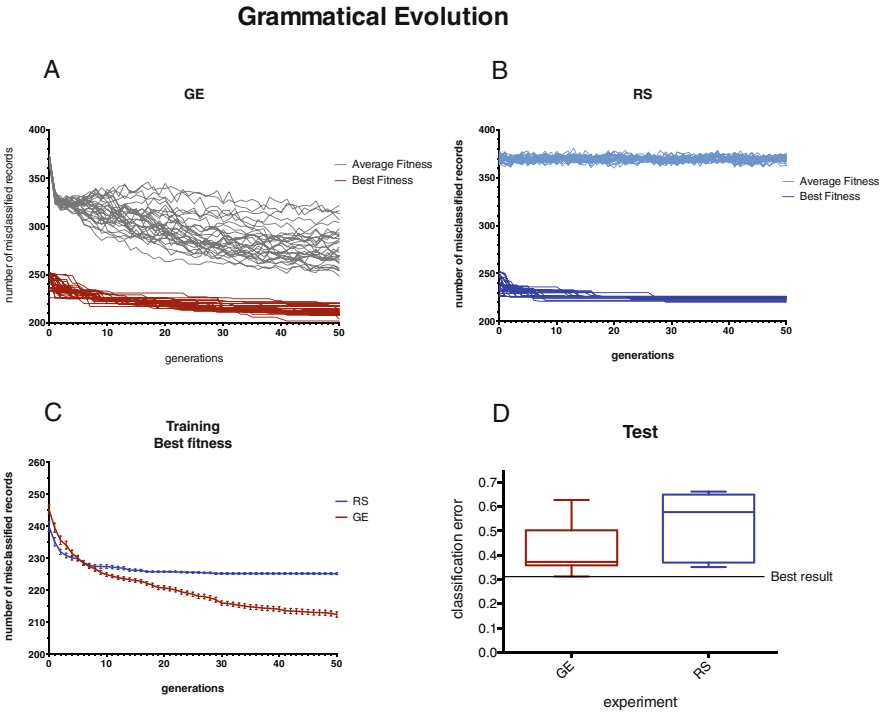

Fig. 2. Training and testing performance in Grammatical Evolution classification experiment. (A) Best and average fitness achieved by GE in each independent run over 50 generations during training. (B) - Best and average fitness achieved by Random Search (RS) in each independent run repeated for 50 generations during training. (C) Mean best fitness achieved by GE and RS during training, (mean sem). (D) Fitness (classification error) returned by evolved best-of-run evolutionary classifiers applied to the test data.

```
A:
[if   (or   x41   (equals   x37   3) )
[if   (and   (equals   x36   36 )  (not_equals   x36   24) )  1   0]
[if   (nor   (greater   (-   x13   -0.2   *   x25)   0.6)  ( greater   -1   0.0))   0
[if   (or   (nand   x39   (xor   x43   x41))   (nor   x39   x42) )  1   0]]]

B:
[if  (or   x41   (equals   x36   23))   0]
[if  (less_e   (+   -0.3   0.4)   x34)
[if  ( less   (+   (x14   *   0.4)   (x7   *   x6) )  x4 )  1   0]   0]
```

Fig. 3. Phenotypes of the best evolved GE classifiers. A: phenotype of the individual with the best training fitness, B: phenotype of the individual with the best fitness on the test dataset.

of these individuals. It can be observed that both classifiers use only 7 predictive variables out of the available 43 variables and both of them use the same 2 variables at the early splits. To provide a benchmark for the results obtained

by GE, we compared them with the results obtained on the same dataset split (training and testing subsets), using 6 different machine learning methods. The models were trained and their parameters tuned using the `train` function from R CARET package. The `trainControl` function was used to specify 10-fold cross-validation as the re-sampling method. Table 3 provides details of the models and optimum parameters that were chosen during training. These settings were used to estimate model performance on the training and the test set. The levels of classification accuracy obtained by different models on the training and the previously unseen test set are presented in Fig. 4. All models, except NSC performed better than No Information Rate (NIR, 0.65) during training. As should be expected, performance of all models on the test set was lower than during training (NIR of test set 0.6308). All models struggled to achieve reliable classification of the test set. The best overall accuracy was achieved by SVM-RBF model (0.663), while the GE model selected on the basis of the best training fitness showed low generalisation to the test set (0.629). It is worth noting that GE was able to evolve another classifier that achieved much higher prediction accuracy on the unseen data (0.688), however performance of this classifier during training (as described above) prevents from the direct comparison of its results with the other models.

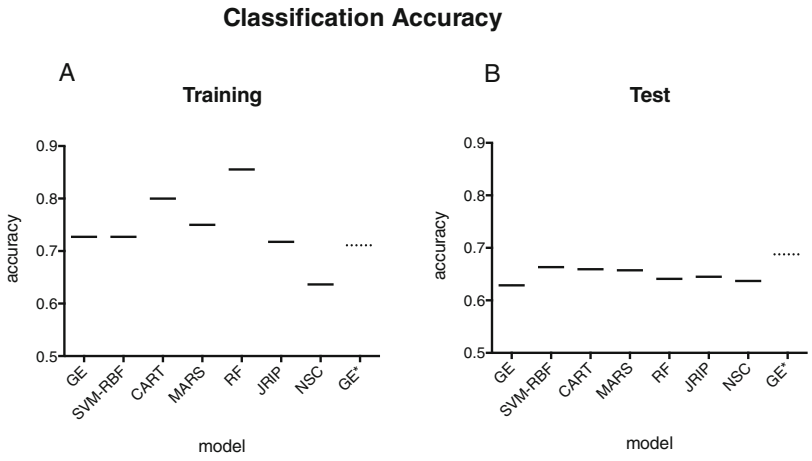

Fig. 4. Comparison of classification accuracy between different models. (A) Best performance during training. (B) Prediction on the test data. GE Grammatical Evolution, SVM Support Vector Machine with RBF (Gaussian) kernel, CART Classification And Regression Tree, MARS Multivariate Adaptive Regression Splines, JRIP One Rule, RF Random Forest, NSC Nearest Shrunken Centroids, GE* - Grammatical Evolution, best in test

There is no one universal best way to compare performance of different models, especially in the case of class imbalances [25]. Our dataset has some class imbalance (36/64 %) with our main class of interest being the minority class.

Classification Accuracy might not be a sufficient measure of model performance in this case. Table 4 gives detailed classification quality metrics obtained by models on the test set. Though SVM_RBF model delivers the best test accuracy and precision, MARS model achieves the best Kappa, while NSC model has the best Sensitivity, F-score and J-statistic, reflecting the superior ability of this model to correctly classify more instances of class 1, but this sensitivity comes at the cost of the ability to correctly classify instances of the second class and consequently, the overall accuracy of the model suffers. The results also show that the second examined GE classifier performs very well, in fact having the best accuracy, precision, Kappa and J-statistic, though does not reach the sensitivity and F-score of the NSC model. In conclusion, results captured in Table 4 highlight that ranking of the models - performance as classifiers greatly depend on the metric chosen, which in turn is dependent on the purpose of the classification (do we place an equal importance on the correct prediction of both classes, or are we specifically interested in correctly predicting instances of one of the classes even if it slightly degrades overall performance).

Table 4. Test Classification Quality Metrics (best result for each metric in bold, * - GE best in test results)

Model	Accuracy (training)	Accuracy (test)	Sensitivity (test)	Specificity (test)	Precision (test)	Kappa (test)	F-score (test)	J-stat (test)
GE	0.727	0.629	0.016	**0.987**	0.429	0.005	0.032	0.004
SVM_RBF	0.727	**0.663**	0.385	0.826	**0.565**	0.226	0.458	0.211
CART	0.800	0.659	0.412	0.804	0.551	0.228	0.472	0.216
MARS	0.750	0.657	0.445	0.781	0.544	**0.235**	0.489	0.227
JRIP	0.718	0.645	0.330	0.830	0.531	0.173	0.407	0.159
RF	**0.855**	0.641	0.456	0.749	0.516	0.210	0.484	0.205
NSC	0.637	0.637	**0.528**	0.701	0.508	0.227	**0.518**	**0.229**
GE*	0.711	**0.688***	0.379	0.868	**0.627***	**0.269***	0.473	**0.247***

We also briefly examined variable importance in different models. Both examined GE classifiers assigned investigators/sites to different classes based on 7 predictive variables. This is a considerably smaller number of variables compared to the lists of variables counted as important by the other top models, though the direct comparison is complicated by the way different models use categorical and numerical variables. Nevertheless, at least 4 of the variables highlighted by GE models appear in the list of variables with above 50 importance score of SVM-RBF, MARS and CART models. This fact further confirms soundness of the evolved GE classifiers.

5 Conclusions and Future Work

This paper adopted a classification approach to predict future performance of investigators/clinical sites in patient recruitment for clinical trials.

The developed GE-based method for classification was compared with a range of well-established machine learning algorithms of different complexity. Taken all together, results are very encouraging, suggesting that GE, in principle, is able to successfully evolve classification model in this scenario. The dataset, at least in its current raw form, proved to be challenging and neither studied model attained a high degree of classification accuracy. The GE classifier taken for the formal comparison with other machine learning algorithms (selected based on the best training performance) showed poor ability to generalise to the test set. However, results of the experiment assessing performance of all 30 evolved GE classifiers show that GE is able to evolve a classifier that is on par or even better than other models. In this first attempt at evolving GE classifier for this business problem we do not have any formal grounds to select the GE classifier that can show better generalisation, though we know that it exists in principle. Based on the observed results and the comparison of the phenotypes of the two examined GE classifiers we can speculate that the best classifier based on training is overfitted to the training set and does not generalise well as the result. This is not surprising, as we have chosen to use decision tree-like structures for our GE classifier and decision trees are known to tend to overfit [14]. Many successful classification algorithms based on decision trees use different techniques to try and avoid such overfitting, such as pruning or early stopping [14]. The issue of overfitting in GE and the generalisation of evolved GE solutions to unseen data were previously highlighted by [21], who examined early stopping as one of the possible approaches to these problems. [24] used monitoring of the generalisation performance of the best-of-generation individual to counteract model overtraining. In the future work we will explore avenues for incorporating early stopping or pruning into our GE set up to avoid overfitting. One of the advantages in the use of GE is a human-interpretable solution [18–20]. In this particular case if we allow ourselves to consider the phenotype of the GE classifier that returned the best results on the test set, we can see that it is not only human-interpretable, but has used only a fraction of predictor variables (7 out of 43) thus effectively performing feature selection simultaneously with classification. So, in depth investigations of these variables might prove to give some additional insights into the factors influencing investigator/site's success in patient enrolment. This study is the first attempt at evolving a GE classifier for this business problem. In this instance we have intentionally used the raw dataset without any data pre-processing in order to establish a baseline for GE model performance. In the future we plan to apply data pre-processing prior to analysis to check whether it will enhance model performance. We also believe that there might be a scope for further increase in the model accuracy through tailored grammar modifications. We also plan to test the use of different fitness functions in order to explore the avenues to tailor the classifier to the business need. The best of the evolved models can be used to help to screen out investigators/sites that have propensity to underperform and jeopardise the trial. Such screening at the early stages of the trial set up can facilitate clinical trial success and substantially reduce costs associated with the need to initiate and maintain

low-performing sites and to bring in "rescue" sites later in the trial. The results clearly demonstrate that GE has the capacity to evolve classification models in this business domain.

Acknowledgments. The authors would like to thank Thomas O'Leary, Pamela Howard and Wilhelm Muehlhausen from ICON Plc. for critical reading of the manuscript and expert advice on patient recruitment and Dr. David Fagan, Dr. Alexandros Agapitos and Stefan Forstenlechner from the UCD Natural Computing Research and Applications Group for their insightful advice on GE methodology. This research is based upon work supported by ICON plc.

References

1. Schueler, P., Buckley, B. (eds.): Re-Engineering Clinical Trials. Best Practices for Streamlining the Development Process, 1st edn. Academic Press Elsevier, Amsterdam (2014)
2. Marks, L., Power, E.: Using technology to address recruitment issues in the clinical trial process. Trends Biotechnol. **20**(3), 105–109 (2002)
3. Trizna, C.: Chapter 9 - no patients, no data: patient recruitment in the 21st century. In: Re-Engineering Clinical Trials. Best Practices for Streamlining the Development Process, 1st edn, pp. 91–105. Academic Press Elsevier (2014)
4. Tufts: CSDD impact report - 89% of trials meet enrolment, but timelines slip, half of sites under-enrol. 15(1) (2013)
5. Kasenda, B., von Elm, E., You, J., Blumle, A., Tomonaga, Y., Saccilotto, R., Amstutz, A., Bengough, T., Meerpohl, J.J., Stegert, M., Tikkinen, K.A.O., Neumann, I., Carrasco-Labra, A., Faulhaber, M., Mulla, S.M., Mertz, D., Akl, E.A., Bassler, D., Busse, J.W., Ferreira-Gonzalez, I., Lamontagne, F., Nordmann, A., Gloy, V., Raatz, H., Moja, L., Rosenthal, R., Ebrahim, S., Schandelmaier, S., Xin, S., Vandvik, P.O., Johnston, B.C., Walter, M.A., Burnand, B., Schwenkglenks, M., Hemkens, L.G., Bucher, H.C., Guyatt, G.H., Briel, M.: Prevalence, characteristics, and publication of discontinued randomized trials. JAMA **311**, 1045–1052 (2014)
6. O'Neill, M., Ryan, C.: Grammatical Evolution: Evolutionary Automatic Programming in a Arbitrary Language, Genetic programming, vol. 4. Kluwer Academic Publishers, Dordrecht (2003)
7. Dempsey, I., O'Neill, M., Brabazon, A.: Foundations in Grammatical Evolution for Dynamic Environments. SCI, vol. 194. Springer, Heidelberg (2009)
8. McKay, R.I., Hoai, N.X., Whigham, P.A., Shan, Y., O'Neill, M.: Grammar-based genetic programming: a survey. Genet. Program. Evolvable Mach. **11**(3/4), 365–396 (2010). Tenth Anniversary Issue: Progress in Genetic Programming and Evolvable Machines
9. Anisimov, V.V., Fedorov, V.V.: Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. Stat. Med. **26**(27), 4958–4975 (2007)
10. Aegerter, P., Bendersky, N., Tran, T.C., Ropers, J., Taright, N., Chatellier, G.: The use of drg for identifying clinical trials centers with high recruitment potential: a feasibility study. Stud. Health Technol. Inf. **205**, 783–787 (2014)
11. Kopcke, F., Lubgan, D., Fietkau, R., Scholler, A., Nau, C., Sturzl, M., Croner, R., Prokosch, H.U., Toddenroth, D.: Evaluating predictive modeling algorithms to assess patient eligibility for clinical trials from routine data. BMC Medical Informatics and Decision Making (2013)

12. Kopcke, F., Prokosch, H.U.: Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J. Med. Internet Res. **16**(7), 161 (2014)
13. Barnard, K.D., Dent, L., Cook, A.: A systematic review of models to predict recruitment to multicentre clinical trials. BMC Medical Research Methodology 10(63) (2010)
14. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. MORGAN KAUFMANN, San Francisco (2011)
15. Koza, J.R.: Hierarchical genetic algorithms operating on populations of computer programs. In: Sridharan, N.S. (ed.) Proceedings of the Eleventh International Joint Conference on Artificial Intelligence IJCAI 1989, vol. 1, pp. 768–774. Detroit, MI, USA, Morgan Kaufmann, 20–25 August 1989
16. Koza, J.R., Keane, M.A., Streeter, M.J., Mydlowec, W., Yu, J., Lanza, G.: Genetic Programming IV: Routine Human-Competitive Machine Intelligence. Kluwer Academic Publishers, Dordrecht (2003)
17. Espejo, P.G., Ventura, S., Herrera, F.: A survey on the application of genetic programming to classification. IEEE Trans. Syst. Man Cybernetics, Part C: Appl. Rev. **40**(2), 121–144 (2010)
18. Nicolau, M., Saunders, M., O'Neill, M., Osborne, B., Brabazon, A.: Evolving interpolating models of net ecosystem CO_2 exchange using grammatical evolution. In: Moraglio, A., Silva, S., Krawiec, K., Machado, P., Cotta, C. (eds.) EuroGP 2012. LNCS, vol. 7244, pp. 134–145. Springer, Heidelberg (2012)
19. Brabazon, A., O'Neill, M.: Diagnosing corporate stability using grammatical evolution. Int. J. Appl. Math. Comput. Sci. **14**(3), 363–374 (2004)
20. Brabazon, A., O'Neill, M.: Credit classification using grammatical evolution. Informatica **30**(3), 325–335 (2006)
21. Tuite, C., Agapitos, A., O'Neill, M., Brabazon, A.: A preliminary investigation of overfitting in evolutionary driven model induction: implications for financial modelling. In: Di Chio, C., et al. (eds.) EvoApplications 2011, Part II. LNCS, vol. 6625, pp. 120–130. Springer, Heidelberg (2011)
22. Kuhn, M.: Building predictive models in r using the caret package. J. Stat. Softw. **28**(5), 1–26 (2008)
23. Ryan, C., Azad, R.M.A.: Sensible initialisation in grammatical evolution. In: Barry, A.M. (ed.) GECCO 2003: Proceedings of the Bird of a Feather Workshops, Genetic and Evolutionary Computation Conference, AAAI 142–145 (2003)
24. Agapitos, A., O'Neill, M., Brabazon, A.: Evolving seasonal forecasting models with genetic programming in the context of pricing weather-derivatives. In: Di Chio, C., et al. (eds.) EvoApplications 2012. LNCS, vol. 7248, pp. 135–144. Springer, Heidelberg (2012)
25. Kuhn, M., Johnson, K.: Applied Predictive Modeling. Springer, New York (2013)