

Comprehensible Models for Predicting Molecular Interaction with Heart-Regulating Genes

Cecilia Sönströd^{*1}, Ulf Johansson¹, Ulf Norinder², Henrik Boström³

¹*School of Business and Informatics, University of Borås, Sweden*

²*AstraZeneca R&D, Södertälje, Sweden*

³*School of Humanities and Informatics, University of Skövde, Sweden*

cecilia.sonstrod@hb.se, ulf.johansson@hb.se,

ulf.norinder@astrazeneca.com, henrik.bostrom@his.se

Abstract

When using machine learning for in silico modeling, the goal is normally to obtain highly accurate predictive models. Often, however, models should also bring insights into interesting relationships in the domain. It is then desirable that machine learning techniques have the ability to obtain small and transparent models, where the user can control the tradeoff between accuracy, comprehensibility and coverage. In this study, three different decision list algorithms are evaluated on a data set concerning the interaction of molecules with a human gene that regulates heart functioning (hERG). The results show that decision list algorithms can obtain predictive performance not far from the state-of-the-art method random forests, but also that algorithms focusing on accuracy alone may produce complex decision lists that are very hard to interpret. The experiments also show that by sacrificing accuracy only to a limited degree, comprehensibility (measured as both model size and classification complexity) can be improved remarkably.

1. Introduction

One of the most intensive areas of research within the pharmaceutical industry is to collect and analyse data on absorption, distribution, metabolism, excretion and toxicity (ADMET) [1]. This is done in order to learn how various compounds interact with the human body, where the main purpose is to guide drug development projects in the search for promising compounds as well as to detect, as early as possible, compounds that should not be put forward as drug candidates (e.g., due to toxicity). A commonly adopted approach is to maintain large series of libraries of chemicals and use high-throughput screening (HTS) to test for biological activity. Promising

compounds found in this way become the focus for continued research, typically leading to continued synthesis and HTS. Synthesis and HTS processes are, however, often time consuming and costly, making it desirable to estimate the biological activity as well as ADMET properties even before synthesis. When computers are used for this initial modeling, the procedure is referred to as *in silico* modeling. If successful, in silico modeling saves much time and investments by excluding unsuitable compounds, allowing earlier focus on promising candidates.

One particular consideration, which is the focus of this paper, concerns avoiding interactions with hERG (human Ether-a-go-go-Related Gene) as well as potential cytochrome P450 interactions related to avoidance of Phase I metabolism, e.g. the nonsynthetic production of more polar compounds (metabolites) of the original chemicals through oxidation, reduction, and hydrolysis. Some drug-induced sudden deaths are a consequence of the development of a side effect of action called *Torsades de Pointes* (TdP). This condition is today a major safety concern for the pharmaceutical industry and health regulatory authorities around the world, see e.g. [2][3][4]. The development of acquired long QT syndrome, which is characterized by the prolongation of the QT interval as a consequence of abnormal behaviour of the cardiac muscle repolarization, is mediated in part by the blockage of the voltage-dependent potassium ion channel encoded by hERG [5]. A number of mathematical models aimed at explaining the relationship between activity for the hERG ion channel and molecular structure have been devised, developed and published in the literature during the last five years. These models are based on statistical techniques such as support vector machines [6][7], neural networks [8], multiple linear regression [9][10][11], as well as recursive partitioning [12][13], and partial least squares analysis [14].

* All four authors are equal contributors to this paper. Corresponding author is Cecilia Sönströd.

As can be seen from the above examples, most machine learning approaches used on this problem are predictive techniques producing opaque models. In these cases, models must have high predictive accuracy to ensure that they can be used efficiently in the drug discovery process. Opaque models, however, have the drawback of not being comprehensible to humans, i.e., they can only be used as black-box prediction machines.

If, on the other hand, models are comprehensible, domain experts can interact with the data mining results in several ways. Computational chemists can not only gain a better understanding of the domain, but also base further analysis and search on the descriptions found. One example is that a comprehensible model might point to interesting sub-domains, where a more targeted search for promising compounds can be conducted using powerful predictive techniques like artificial neural networks, support vector machines or some ensemble technique.

When viewed from a machine learning perspective, the problem thus has characteristics from both predictive and descriptive tasks, as described in [15]. The task resembles predictive modeling, since the objective is to predict the value of a target variable based on a number of attributes. Since computational chemists also seek explanations of underlying relationships in data, the problem is at the same time descriptive in nature. Models must then combine the key properties of relatively high accuracy and comprehensibility. High accuracy is needed in order to ensure that models capture relationships that are general and not only true for the particular observations from which the model was built, and comprehensibility is essential if domain experts are to gain insights about the underlying domain from the model. Since comprehensible models are not primarily intended to be used for prediction, they really do not need to describe the entire data set. In fact, the key demand is that the relationships that the user is interested in are as accurate as possible. Thus, when evaluating models for these kinds of problems, both model accuracy and accuracy for partial descriptions (e.g. single rules) are relevant.

With the above problem, and other similar ones in mind, there is clearly a need for machine learning techniques that are capable of producing transparent, small and yet accurate models. The overall purpose of this paper is to argue for the need for such techniques, and to discuss some key properties. In addition, we compare two standard techniques for generating decision lists, RIPPER and PART, to a technique specifically tailored to meet these needs, using a data set from the drug discovery domain.

2. Background

It should be noted that supervised machine learning techniques, whether producing opaque or transparent

models, normally focus solely on maximizing accuracy. Consequently, even transparent models are quite seldom easily interpretable. A striking example of this is a decision tree containing hundreds of nodes, where each classification is typically made after more than 10 tests, many branches are quite similar and a relatively small number of instances reach each leaf.

So, the fact that most high-accuracy models are outright opaque, while most transparent models are only of limited comprehensibility due to their complexity, the need for techniques tailored to produce compact, yet accurate models, is apparent. A domain expert that aims to understand the reasons for a particular prediction, or seeking to gain novel insights, is often not interested in the entire model, but rather parts of it, i.e. specific rules. Hence, it is often not the size of the entire model (e.g. as measured by the number of rules or tests) that is the most important factor for comprehensibility, but rather the complexity of parts of the model (e.g. rules) that are employed when making a specific prediction. The measure classification complexity (CC), introduced in [16], is one alternative complexity measure that is motivated by this fact. Having chosen a suitable measure for comprehensibility (complexity), a desirable feature of any learning algorithm that aims for achieving both high accuracy and high comprehensibility is to allow balancing these factors, e.g. by means of a parameter, depending on the requirements of the application.

In this work, we focus on methods for learning ordered rule sets (or decision lists) as a way of obtaining compact, and yet accurate, models. Obviously, many techniques for producing decision lists exist; early examples of decision list algorithms include AQ [17] and CN2 [18]. More recent is RIPPER [19], based on IREP [20]. These are all based on the sequential covering algorithm, where the central idea is that rules are learnt one at a time. For each rule, all instances covered by this rule are removed from the data set and the next rule is learnt from the remaining instances. Of these techniques, RIPPER is regarded as being the state of the art [21], and it includes a powerful optimization procedure. However, RIPPER cannot be considered entirely suitable for the above purpose, since it does not contain any obvious way of balancing accuracy against comprehensibility. The standard RIPPER version also has the disadvantage of only formulating rules for the minority class, which means that it will not provide users with clear descriptions of all relationships in the data set. Another property that sometimes make RIPPER rule sets hard to interpret is the use of conjunctive rules; a typical RIPPER rule consists of several (in general between 3 and 7) conjuncts. Another technique for generating decision lists is PART, which is implemented in Weka [21], where rules are obtained from partial decision trees built using the same heuristic as C4.5 trees [22].

In previous studies [23][24], different ways of handling problems similar to the present one have been investigated from a data mining perspective, where such problems were identified as being instances of the data mining task *concept description* as defined in CRISP-DM [25]. CRISP-DM states that the aim of concept description “is not to develop complete models with high prediction accuracy, but to gain insights”. In [16], the Chipper algorithm, which is specifically designed for concept description was introduced and evaluated.

Chipper is a deterministic algorithm for generating decision lists consisting of simple rules. In its current implementation, it handles only binary classification tasks. The basic idea is to, in every step, search for the rule that covers the maximum number of instances, with sufficient accuracy, using a split on one attribute. For continuous attributes, this means a single comparison using a relational operator, (\leq , \geq), and for nominal attributes a set of instances having identical values for that attribute ($=$).

Two main parameters, called *ignore* and *stop*, are used to control the rule generation process. The *ignore* parameter specifies the misclassification rate that is acceptable for each rule and can have different values for each output class. The *ignore* parameter is given as a percentage of remaining instances or as an absolute number of instances. The motivation for the *ignore* parameter is that it can be used to view the data set at different levels of detail, with higher values prioritizing the really broad discriminating features of data items and low values trying to capture more specific rules. The *stop* parameter specifies the proportion of instances in the data set that should be covered by rules before terminating. The motivation for this parameter is that it can be used to find only the most general relationships in the data, instead of trying to find rules to cover particular instances. In effect, these two parameters control the level of “granularity” for the decision list.

Experimentation on UCI [26] data sets in [16] has shown that Chipper obtains similar accuracies to both the standard decision tree algorithm C4.5, implemented as J48 in Weka and RIPPER, implemented in Weka as JRip.

3. Method

The aim of the experiments is to investigate whether the evaluated decision list algorithms (Chipper, JRip and PART) can produce comprehensible models of relationships in the hERG data set that are of use to computational chemists. For comparison regarding accuracy level and comprehensibility, the data set was first analyzed using random forests as implemented in the system RDS [27]. Comprehensibility was measured using rule set size; both number of rules and number of tests, since some techniques use conjunctive rules, and

classification complexity (CC). Classification complexity measures the average number of rules or tests that are needed to classify an instance. When calculating *CC test* for a conjunctive rule, the rule is taken to classify all instances at the last conjunct, since all conditions are needed to obtain a classification.

Finally, the effects of the mechanisms for controlling accuracy, comprehensibility and coverage in Chipper were evaluated.

3.1. Data set

The data set consisted of 6020 drug-like molecules from the AstraZeneca corporate database. These compounds have been screened for hERG blockade using the IonWorks™ HT (high throughput) electrophysiology assay [28][29]. The response from the assay was subsequently divided into two classes (high and low). The molecules were transformed into their respective neutral states, e.g. no overall charge was assigned to the compounds. Molecular properties (descriptors) were generated using the in-house program SELMA [30]. SELMA generates parameters related to size, flexibility, hydrogen bonds, electronics (charge related properties), as well as lipophilicity. In total, SELMA computed 196 parameters describing the molecules in the data set.

3.2. Techniques

Decision trees have many attractive features, such as allowing for human interpretation and hence making it possible for a decision maker to gain insights into what factors are important for particular classifications. Recent research has shown that significant improvements in predictive performance can be achieved by generating large sets of models, or *ensembles*, which are used to form a collective vote on the value for the dependent variable [31]. For ensemble models to work as intended there must be some diversity among the base classifiers. One popular method of introducing diversity is *bootstrap aggregating*, or *bagging*, as introduced by Breiman [32]. Another popular method of introducing diversity when generating decision trees is to consider only a small subset of all available independent variables at each node when forming the tree. When combined with bagging, the resulting models are referred to as *random forests* [33], and these are widely considered to be among the most competitive and robust of current methods for predictive data mining. The drawback of ensemble models is, however, that they can no longer be easily interpreted and hence provide less guidance into how classifications are made.

The Rule Discovery System (RDS) [27] addresses this problem by providing some insight into what factors are of importance in an ensemble of decision trees by

presenting the *variable importance* of each independent variable, i.e. how much the variable, relative to all other variables, contributes to reducing the squared error of the dependent variable.

3.3. Experiments

A two-fold cross-validation procedure was employed throughout the work. In Experiment 1, the data set was initially analyzed by ensemble modeling (50 trees) and recursive partitioning using RDS™. Variable selection was then performed based on the results from the ensemble modeling, where the 90 most important parameters were subsequently used for further analysis in Experiment 2 using Chipper, JRip and PART.

Experiment 3 was designed to investigate the tradeoff between accuracy, model size and coverage, by varying the *ignore* and *stop* settings in Chipper. In all, 14 Chipper runs were performed, consisting of all possible combinations of the following parameter settings:

- Stop: 70, 90
- Ignore: 0.25%, 0.5%, 1%, 2%, 4%, 8%, 16%

4. Results

In the first experiment, the data set was analyzed using RDS, followed by variable selection. Reduction of the number of parameters from 196 to 90 (54% reduction) did not influence the accuracy of the derived model, nor the number of rules generated to any significant extent. Table 1 below shows accuracy and size results for RDS, using both the full and the reduced data set.

Table 1. RDS Accuracy and size

	Full		Reduced	
	Acc.	#rules	Acc.	#rules
	RDS tree	69.3	102	70.3
RDS ensemble	80.8	12160	79.9	11256

As can be seen, there is a substantial difference between single tree and ensemble accuracy, which is in accordance with established ensemble theory. The ensemble is, however, inherently opaque and thus not comprehensible. The single tree model, although transparent, contains over 100 rules and thus holds limited explanatory value.

In light of the discussion above regarding accuracy and comprehensibility in decision lists, it is of interest to take a more detailed look at the decision lists produced. Table 2 below shows these results for the reduced data set. The column *r_acc* (rule accuracy) shows the accuracy of the rules; i.e. without the default rule, *m_acc* (model accuracy) on the other hand includes instances classified by the default rule. For this experiment, Chipper was set at 4.5% *ignore* and the *stop* parameter was set at two different values (70% and 90%).

Table 2. Accuracy, coverage and size

	<i>m_acc</i>	<i>r_acc</i>	# rules	# tests	CC rule	CC test
PART	76.4	91.3	169	996	98.5	411.1
JRip	74.2	82.2	20	93	16.1	74.7
Chip90	67.2	75.1	14	14	5.2	5.2
Chip70	67.4	76.4	7	7	4.0	4.0

The main result from Experiment 2 is the strong relationship between accuracy and rule size that holds for all three decision list algorithms. This indicates that relationships in the data set are complex and that it takes many conditions to build a model with high accuracy.

Regarding classification complexity, PART performs very badly, requiring on average 100 rules, containing altogether several hundred conditions, to classify an instance. JRip performs better, but still needs on average 75 conditions to classify an instance, mainly due to only describing the minority class. Chipper, finally, manages to use its small rule sets rather efficiently and obtains very low classification complexities.

The results from Experiment 3, regarding Chipper and the effect of different parameter settings on accuracy and comprehensibility, are shown in Table 3 below. The *r_acc train* and *r_acc test* columns show the average training and test accuracy over the two folds, respectively. *#rules* is the average number of rules over the two folds. Note that only rule accuracy is shown in the table; model accuracy is however, stable around 68%.

Table 3. Chipper size and accuracy

Stop Ign	70			90		
	<i>r_acc train</i>	<i>r_acc test</i>	# rules	<i>r_acc train</i>	<i>r_acc test</i>	# rules
0.25%	0.872	0.715	83	0.879	0.691	158.5
0.5%	0.846	0.715	46.5	0.846	0.693	89.5
1%	0.820	0.736	25	0.812	0.697	52
2%	0.814	0.756	13	0.799	0.731	27.5
4%	0.783	0.747	7	0.770	0.713	14.5
8%	0.744	0.700	4	0.738	0.690	7.5
16%	0.720	0.712	2	0.719	0.703	3

As can be seen in the table, rule set size, as expected, correlates very well with the *ignore* value; increasing *ignore* invariably leads to a smaller rule set. Regarding rule test accuracy, the pattern is less clear, but results seem to indicate that an *ignore* value in the middle range gives higher rule set accuracy. Unsurprisingly, low *ignore* values lead to over-fitting, with training accuracies approaching 90%, but much worse rule accuracy. Using a too high *ignore* value, on the other hand, lowers training rule accuracy considerably. This pattern holds for both settings of the *stop* parameter. The *ignore* parameter thus works as intended, allowing the user to select the level of

detail with which to describe relationships. For this particular, quite difficult data set, the effect on training accuracy is immediate, but this does not carry over to test accuracy.

Comparison between the two *stop* values shows that lowering the *stop* value, i.e. reducing the coverage, will decrease model size in a predictable way. Using a high *stop* value also has the disadvantage of increasing the risk of over-fitting, as seen by the worse test accuracies for all *ignore* values.

In Figure 1 below, a sample Chipper rule set, for 3% *ignore* and 70% *stop*, is shown:

```
R1:  HYBOT_sum_donor  >= -1.279  -> high 1079/180
R2:  GClogP          <=  1.911  -> low  661/148
R3:  PSA             <= 51.891  -> high 487/126
R4:  HYBOT_max_donor >= -1.660  -> high 441/92
R5:  GC-type-058    <= -0.003  -> high 360/65
R6:  Neg_ioniz.     >=  1.001  -> low  319/72
R7:  GC-type-026    >=  4.994  -> high 283/79
R8:  Motoc_index    >=  9.135  -> high 235/71
R9:  Mol._volume_2D >= 846.72  -> high 186/63
R10: Motoc_index    >=  8.602  -> high 185/59
```

Figure 1. Sample Chipper rule set

The above rule set, as derived by Chipper with parameters at 70% *stop* and 3% *ignore*, consisted of 10 rules. 778 compounds (~13 %) were not covered by the derived rules, which is far below the limit of 30%. Average rule accuracy over the two folds was 74.6% and average model accuracy was 67.8 %.

The properties covered by the rules are consistent with other publications indicating similar properties to be important for high and low hERG activity, respectively.

Thus, lipophilic, unpolar and large compounds are the type of compounds that would interact strongly with hERG, as exemplified by rules R1-R4, R6 and R9. In rule R1 and R4, the lack of hydrogen bonds, i.e. a polar characteristic, results in hERG active compounds (note that the HYBOT donor variables are negatively defined, i.e., the more hydrogen bonds, the larger negative value and, consequently, the more positive value the less hydrogen bonding properties, and the more unpolar character of the molecule). In rule R2, it is indicated that a lipophilic (fat) compound, i.e. a high value for the GClogP variable, also gives hERG active compounds. In rule R6, it is indicated that negatively ionisable compounds, e.g. carboxylic acids, are not hERG active which is also in agreement with present knowledge.

This short set of interpretable rules enables researchers to focus their attention on compounds with suitable molecular properties that are likely to be devoid of hERG activity.

5. Conclusions

In this study, the task of predicting hERG activity from a large number of attributes has been studied from a

machine learning perspective. For this problem, it is desirable to obtain comprehensible models, in order to gain insights into interesting relationships in the underlying domain. Various demands on such models and techniques used to obtain them have been discussed. Most importantly, techniques should always be able to produce comprehensible models, and be able to obtain reasonable accuracy, given this restriction.

The results from the experiments show that algorithms for learning decision lists can obtain predictive performance that is not far from the state-of-the-art method random forests. However, decision list algorithms focusing on accuracy alone may give rise to very complex decision lists that can be very hard to interpret, which is in direct contrast to the purpose of generating interpretable models in the first place. The experiments show that by sacrificing accuracy only to a limited degree, the comprehensibility (measured both by size of the rule sets and by classification complexity) can be improved remarkably. This, in turn, gives interpretable information to researchers and enables them to focus their attention on more suitable sub-domains to avoid interactions with hERG.

Acknowledgment

This work was supported by the Information Fusion Research Program (www.infofusion.se) at the University of Skövde, Sweden, in partnership with the Swedish Knowledge Foundation under grant 2003/0104.

References

- [1] H. van de Waterbeemd and E. Gifford, "ADMET in silico modelling: towards prediction paradise?", *Nat Rev Drug Discov.*, Vol. 2:3: pp. 192-204, 2003.
- [2] A.M. Brown, "Drugs, hERG and sudden death", *Cell Calcium*, 35(6), pp. 543-547, 2004.
- [3] R.R. Shah, "The significance of QT interval in drug development", *Br J Clin Pharmacol*, 54(2) pp. 188-202, 2002.
- [4] R.R. Fenichel, M. Malik, C. Antzelevitch, M. Sanguinetti, D.M. Roden, S.G. Priori, J.N. Ruskin, R.J. Lipicky, L.R. Cantilena, "Drug-induced torsades de pointes and implications for drug development", *J Cardiovasc Electrophysiol*, 15(4) pp. 475-495, 2004.
- [5] M.C. Sanguinetti, C. Jiang, M.E. Curran, M.T. Keating, "A mechanistic link between an inherited and an acquired cardiac arrhythmia: hERG encoded the I_{Kr} potassium channel", *Cell. Physiol. Biochem.*, 81 pp. 299-307, 1995.
- [6] M. Song and M. Clark, "Development and evaluation of an in silico model for hERG binding", *J Chem Inf Model* 46(1), pp. 392-400, 2006.
- [7] M.K. Leong, "A novel approach using pharmacophore ensemble/support vector machine (PhE/SVM) for prediction of hERG liability", *Chem Res Toxicol*, 20(2), pp. 217-226, 2007.
- [8] M. Seierstad and D.K. Agrafiotis, "A QSAR model of hERG binding using a large, diverse, and internally

- consistent training set”, *Chem Biol Drug Des*, 67(4), pp. 284-296, 2006.
- [9] K. Yoshida K and T. Niwa, “Quantitative structure-activity relationship studies on inhibition of hERG potassium channels”, *J Chem Inf Model*, 46(3) pp. 1371-1378, 2006.
- [10] A. Coi, I. Massarelli, L. Murgia, M. Saraceno, V. Calderone, A.M. Bianucci, “Prediction of hERG potassium channel affinity by the CODESSA approach”, *Bioorg Med Chem*, 14(9), pp. 3153-3159, 2006.
- [11] S. R. Johnson, H. Yue, M.L. Conder, H. Shi, A.M. Doweyko, J. Lloyd, P. Levesque, “Estimation of hERG inhibition of drug candidates using multivariate property and pharmacophore SAR”, *Bioorg Med Chem*, 15(18), pp. 6182-6192, 2007.
- [12] M.M. Gepp and M.C. Hutter, “Determination of hERG channel blockers using a decision tree”, *Bioorg Med Chem*, 14(15), pp.5325-5332, 2006.
- [13] E. Dubus, I. Ijjaali, F. Petitet, A. Michel, “In silico classification of hERG channel blockers, A knowledge-based strategy”, *ChemMedChem*, 1(6), pp. 622-630, 2006.
- [14] C.L. Gavaghan, C.H. Arnby, N. Blomberg, G. Strandlund, S. Boyer, “Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data”, *J Comput Aided Mol Des*, 21(4), pp. 189-206, 2007.
- [15] P. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, Pearson Education, 2006.
- [16] U. Johansson, C. Sönströd, T. Löfström and H. Boström, “Chipper – A Novel Algorithm for Concept Description”, *10th Scandinavian Conference on Artificial Intelligence*, IOS Press, pp. 133-140, Stockholm, Sweden, 2008.
- [17] R. S. Michalski, “On the quasi-minimal solution of the general covering problem”, *Proceedings of the Fifth International Symposium on Information Processing*, p.125-128, Bled, Yugoslavia, 1969.
- [18] P. Clark and T. Niblett, “The CN2 induction algorithm”, *Machine Learning*, 3: 261-283, 1989.
- [19] W. Cohen, “Fast Effective Rule Induction”, *Proceeding of 12th International Conference on Machine Learning (ICML '95)*, Tahoe City, CA, p. 115-123, 1995.
- [20] J. Fürnkranz and G. Widmer, “Incremental Reduced Error Pruning”, *11th International Conference on Machine Learning (ICML '94)*, San Mateo, CA, p. 70-77, 1994.
- [21] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques 2nd ed.*, Morgan Kaufman, 2005.
- [22] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufman, 1993.
- [23] U. Johansson, C. Sönströd and L. Niklasson, “Explaining Winning Poker – A Data Mining Approach”, *6th International Conference on Machine Learning and Applications*, Orlando, FL, IEEE press, pp. 129-134, 2006.
- [24] C. Sönströd, U. Johansson, and R. König, “Towards a Unified View on Concept Description”, *The 2007 International Conference on Data Mining (DMIN07)*, Las Vegas, NV, 2007.
- [25] The CRISP-DM Consortium, CRISP-DM 1.0, www.crisp-dm.org, 2000.
- [26] C. J. Asuncion, A. & Newman, D.J., UCI Machine Learning Repository [http://www.ics.uci.edu/~mllearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.
- [27] Rule Discovery System, v. 2.6.0, Compumine AB, http://www.compumine.com/web/public/rds (accessed May 29, 2008)
- [28] K. Schroeder, B. Neagle, D.J. Trezise, J. Worley, “IonWorks™ HT: A New High-Throughput Electrophysiology Measurement Platform”, *J Biomol Screen*, 8, pp. 50-64, 2003.
- [29] M.H. Bridgland-Taylor, A.C. Hargreaves, A. Easter, A. Orme, D.C. Henthorn, M. Ding, A.M. Davis, B.G. Small, C.G. Heapy, N. Abi-Gerges, F. Persson, I. Jacobson, M. Sullivan, N. Albertson, T.G. Hammond, E. Sullivan, J.P. Valentin, C.E. Pollard, “Optimization and validation of a medium-throughput electrophysiology-based hERG assay using IonWorks HT”, *J. Pharm. Tox. Methods*, 54(2), pp. 189-199, 2006.
- [30] T. Olsson and V. Sherbukhin, “SELMA, Synthesis and Structure Administration (SaSA)”, AstraZeneca R&D Mölndal, Sweden.
- [31] E. Bauer and R. Kohavi, “An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants”, *Machine Learning*, vol. 36, pp. 105-139, 1999.
- [32] L. Breiman, “Bagging Predictors”, *Machine Learning*, vol. 24, pp. 123-140, 1996.
- [33] L. Breiman, “Random Forests”, *Machine Learning*, vol. 45, pp. 5-32, 2001.