# Causal Mechanisms and Classification Trees for Predicting Chemical Carcinogens

**Louis Anthony Cox, Jr.**
Cox Associates, tony@cox-associates.com
503 Franklin Street, Denver, CO, 80218

## Abstract

Classification trees, usually used as a nonlinear, nonparametric classification method, can also provide a powerful framework for comparing, assessing, and combining information from different expert systems, by treating their predictions as the independent variables in a classification tree analysis. This paper discusses the applied problem of classifying chemicals as human carcinogens. It shows how classification trees can be used to compare the information provided by ten different carcinogen classification expert systems, construct an improved "hybrid" classification system from them, and identify cost-effective combinations of assays (the inputs to the expert systems) to use in classifying chemicals in future.

## 1 INTRODUCTION

One of the most difficult applications challenges for statistical and AI classification technology has turned out to be predicting which chemicals are likely to cause cancer in humans, without performing costly experiments in mice and rats to find out. Part of the difficulty stems from the fact that the term "carcinogen" applies to chemicals that operate by radically different causal mechanisms to produce very different biological responses involving uncontrolled cell proliferation, all of which are referred to as "cancer" (Williams 1996). Learning the concept "carcinogen" from training data therefore requires learning a disjunction of concepts that are heterogeneous in terms of the physical reality being described -- the relevant chemical structures and physiochemical properties, the biological systems affected, and the spectrum of biological responses produced in test systems. The extension of the term "carcinogen" involves an intrinsically complex and heterogeneous ontology that cannot easily be represented by few or simple relations among attributes in a training database.

Despite this complexity, carcinogenicity in mice and rats often predicts carcinogenicity in humans (Ashby and Paton, 1993). More specifically, chemical carcinogens can usefully be subdivided into *genotoxic* carcinogens, which cause cancer by reacting with DNA, and *non-genotoxic* carcinogens, which involve other causal mechanisms leading to stimulated proliferation of Williams 1996, Chevalier 1998). Strongly genotoxic carcinogens often cause cancer in multiple species, sexes, strains, and organs by a common DNA-damaging mechanism (Gold 1991). Therefore, potent mouse- and rat-carcinogens are often considered to be potential human carcinogens. Inferring likely human carcinogenicity for non-genotoxic chemicals, however, is a largely unsolved problem (Ashby 1993). An example of a non-genotoxic mechanism is found in experiments with diesel exhaust (DE), which can cause lung cancer in rats at high, prolonged exposures by forming soot deposits that repeatedly abrade and irritate the lung tissue, eventually depleting protective enzymes and inducing compensating proliferation of cells. This increased proliferation, in turn, raises the probability that at least one cancerous cell will arise. Such non-genotoxic mechanisms tend to be highly species-specific (Ashby 1993, Chevalier 1998). For example, DE does not appear to cause lung cancer or deplete protective enzymes in other species (Cox, 1997).

Genotoxic chemical carcinogens often have structural similarities (such as a "bay region" in a multi-ring organic molecule) that once seemed promising for predicting carcinogenicity. Yet, non-genotoxic carcinogens constitute a miscellany of chemicals, from simple organics like chloroform (Templin 1998,) to complex ones like DE, that increase cell proliferation by various idiosyncratic mechanisms (Williams 1996, Yoshikawa 1996). This creates an inherently deceptive setting for many machine-learning and automated inference or concept-learning programs. Patterns that might prove predictively useful if only genotoxic chemicals were considered become diluted and confounded by non-genotoxic chemicals. The result is that even relatively sophisticated predictive systems often perform poorly when tested on chemicals for which the correct classification is initially unknown (Benigni 97).

This paper introduces a new approach to predicting chemical carcinogens. It is motivated by the observation that different current predictive systems incorporate some complementary and some redundant information about relevant aspects of chemical structures, properties, and effects in various assays and biological systems. Analyzing the empirical performance (i.e., prediction accuracy and failure patterns) of these different

algorithms leads to a relatively rich model of how their errors are interrelated. This, in turn, reveals how their predictions can best be combined to obtain a hybrid model that out-performs any of the individual models that contribute to it.

## 2 AN ILLUSTRATIVE EXAMPLE

Figure 1 illustrates one such hybrid predictive model, based on the performance data from the ten individual predictive systems summarized in Table 1.

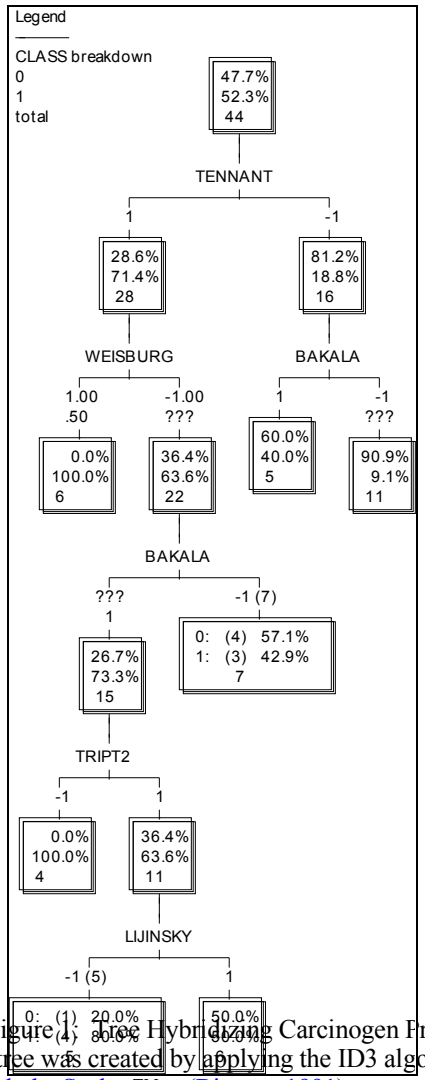


Figure 1. Tree Hybridizing Carcinogen Predictions

This tree was created by applying the ID3 algorithm in KnowledgeSeeker™ (Biggs 1991) to dichotomized outcome data in which each of 44 chemicals was classified as either a rodent carcinogen or as not clearly carcinogenic, based on the outcomes of long-term cancer bioassays in mice and rats. For clarity of exposition, we dichotomized the ternary outcomes used by many prediction systems, which classify chemicals as carcinogens, non-carcinogens, or equivocal/uncertain carcinogens in rodents, e.g., based on whether cancer is predicted to occur in both, neither, or just one of mice and rats. In Figure 1 and subsequent trees, "Class 1" represents a rodent carcinogen in these bioassays, while "Class 0" represents a non-carcinogen or equivocal carcinogen, using Bristol's (1996) summary of bioassay results.

The independent variables entering the classification tree analysis were predictions of carcinogenicity from each of ten individual predictive expert systems (i.e., Tennant, Weisburger, etc.) summarized in Table 1. The predictions are dichotomous or ordered-polytomous. We have represented all predictions by numerical scores, e.g., using the ordinal scale -1 = predicted non-carcinogen, 0.25 = predicted possible carcinogen, 0.5 = predicted probable carcinogen, 1 = predicted carcinogen, and ??? = no prediction (e.g., because data required to make a prediction were missing.)

In Figure 1, the single system that best predicts rodent carcinogenicity of the 44 test chemicals is that of Tennant, as also noted by Benigni (1995). If this system classifies a chemical as a rodent carcinogen, then there is a 71.4% chance (in this sample of 44 chemicals) that the long-term cancer bioassays will indeed show it to be a rodent carcinogen. Conversely, a chemical classified as a non-carcinogen has an 81.2% chance of not being a clear rodent carcinogen when the long-term bioassay is conducted. Thus, the total "resubstitution" error rate for the Tennant predictions alone is 25% = (28/44)(28.6%) + (16/44)(18.8%). Since these predictions are based on structural alerts learned from a training set of chemicals that did not include the 44 test chemicals considered in Figure 1, these estimates of performance are probably realistic. Resubstitution misclassification error rates for other individual predictive systems are as follows:

| | | |
|---|---|---|
| Tennant = 25% | Benigni = 34% | COMPACT = 41% |
| Weisburger = 32% | Bakale = 39% | DEREK = 41% |
| Tript 1 = 34% | CASE = 41% | TOPKAT = 43% |
| RASH = 34% | Tript 2 = 41% | Lijinsky = 45.5% |

Table 1: Predictions and Results of Rodent Carcinogenicity Tests

| NUMCHEM | MR | FR | MM | FM | CLASS | TENNANT | TRIPTI | BENIGNI | WEISBURGER | BAKALE | TOP_KAT | TRIPT | DEREK | COMPACT | LIJINSKY | MLT_CASE | DEREK_HY | RASH |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NE | NE | NE | NE | NEG | -1 | -1 | -1 | -1 | | | 1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 2 | NT | NT | NE | SE | POS | -1 | -1 | -1 | -1 | -1 | | -1 | -1 | 1 | -1 | -1 | -1 | |
| 3 | EE | NE | NE | NE | EQV | -1 | -1 | | -1 | -1 | | -1 | | | -1 | 1 | | 0 |
| 4 | NE | NE | NE | NE | NEG | -1 | -1 | -1 | -1 | -1 | | -1 | -1 | 1 | 1 | -1 | -1 | -1 |
| 5 | NE | NE | NE | NE | NEG | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -0 |
| 6 | NE | NE | EE | NE | EQV | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 |
| 7 | NE | NE | EE | EE | EQV | -1 | -1 | | -1 | | | -1 | -1 | | -1 | | -1 | 1 |
| 8 | NE | NE | NE | NE | NEG | -1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | 1 | -1 |
| 9 | NT | NT | NE | NE | NEG | -1 | -1 | 1 | -1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | 1 | -1 |
| 10 | NE | NE | NE | NE | NEG | -1 | 1 | -1 | -1 | 1 | | 1 | -1 | 1 | 1 | 1 | -1 | -1 |
| 11 | NE | EE | SE | NE | POS | 1 | 1 | 0 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 |
| 12 | CE | CE | CE | CE | POS | 1 | 1 | 0 | 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | |
| 13 | SE | NE | EE | SE | POS | 1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -0 |
| 14 | SE | NE | NE | SE | POS | 1 | 1 | -1 | -1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | 0 |
| 15 | NT | NT | NE | NE | NEG | 1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 16 | SE | EE | EE | NE | POS | 1 | 1 | | -1 | -1 | | 1 | | | -1 | | | 0 |
| 17 | NE | NE | SE | SE | POS | 1 | -1 | -1 | -1 | | -1 | -1 | 1 | -1 | -1 | -1 | 1 | 1 |
| 18 | NE | NE | NE | NE | NEG | 1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 |
| 19 | NE | NE | NE | NE | NEG | 1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | 0 | -1 | 1 | -1 | -1 |
| 20 | EE | NE | SE | SE | POS | 1 | 1 | -1 | -1 | | -1 | 1 | -1 | 1 | 1 | 1 | 1 | -1 |
| 21 | EE | NE | NE | CE | POS | 1 | 1 | 0 | 0 | -1 | | -1 | -1 | -1 | 1 | -1 | -1 | 1 |
| 22 | SE | EE | SE | NE | POS | 1 | 1 | 0 | -1 | 1 | -1 | 1 | -1 | 1 | 1 | 1 | 1 | -1 |
| 23 | NE | EE | NE | NE | EQV | -1 | -1 | 0 | -1 | | | 1 | -1 | 1 | -1 | | -1 | |
| 24 | NE | NE | NE | NE | NEG | -1 | -1 | 0 | -1 | | 1 | -1 | -1 | 1 | -1 | 1 | 1 | |
| 25 | NE | NE | NE | NE | NEG | -1 | 1 | 0 | 0 | 1 | | 1 | 1 | -1 | 1 | 1 | 1 | |
| 26 | NT | NT | NE | NE | POS | -1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 |
| 27 | NE | NE | NT | NT | NEG | -1 | 1 | | -1 | | | 1 | -1 | -1 | -1 | | 1 | -1 |
| 28 | CE | CE | EE | EE | POS | -1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | -1 | 1 | 1 | 1 | -1 |
| 29 | SE | NE | CE | CE | POS | 1 | 1 | 0 | -1 | | | 1 | 1 | 1 | -1 | 1 | 1 | -1 |
| 30 | SE | SE | SE | NE | POS | 1 | 1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 31 | EE | NE | NE | NE | EQV | 1 | 1 | 1 | -1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 32 | NE | NE | SE | NE | POS | 1 | 1 | 0 | -1 | -1 | 1 | 1 | -1 | 1 | 1 | 1 | -1 | 1 |
| 33 | NE | EE | NE | NE | EQV | 1 | 1 | 0 | -1 | -1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 | -0 |
| 34 | SE | SE | CE | CE | POS | 1 | -1 | 1 | | | 1 | -1 | 1 | 1 | 1 | -1 | 1 | |
| 35 | EE | EE | NT | NT | EQV | 1 | 1 | | | 1 | | 1 | -1 | | -1 | | -1 | -1 |
| 36 | CE | CE | NT | NT | POS | 1 | 1 | 1 | | | | 1 | 1 | 1 | -1 | -1 | 1 | |
| 37 | CE | CE | NT | NT | POS | 1 | 1 | 0 | | | 1 | 1 | | 1 | -1 | -1 | | |
| 38 | SE | EE | SE | CE | POS | 1 | 1 | 0 | | 1 | | 1 | -1 | 1 | -1 | -1 | -1 | 1 |
| 39 | CE | CE | CE | CE | POS | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | -1 | 1 | 1 | 1 | 1 |
| 40 | CE | CE | NT | NT | POS | 1 | 1 | 0 | 1 | -1 | 1 | 1 | -1 | 1 | 1 | 1 | 1 | 1 |
| 41 | EE | NE | NE | NE | EQV | 1 | 1 | 0 | -1 | | -1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| 42 | NT | NT | EE | NE | EQV | 1 | 1 | 1 | -1 | 1 | 1 | 1 | -1 | 1 | 1 | 1 | 1 | 0 |
| 43 | CE | CE | CE | SE | POS | 1 | 1 | 1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | 1 | 0 |
| 44 | CE | CE | CE | CE | POS | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | 1 | 1 | 1 | 0 |

*Source:* Bristol 1996

Clearly, no single test has a very low classification error rate, although 25% for the Tennant system is significantly lower than the error rate for other systems.

The classification tree in Figure 1 combines predictions from five different systems. For example, if both Tennant *and* Weisburger predict that a chemical is a rodent carcinogen, then so does the tree that hybridizes them (since 6 out of 6 cases predicted to be carcinogenic by Tennant and classified as carcinogenic or probably carcinogenic by Weisburger were in fact observed to be carcinogenic in the rodent bioassays). When Tennant predicts that a chemical is a rodent carcinogen and Weisburger does not, however, there is about a 64% chance that the chemical is carcinogenic. Introducing results of other tests, such as Bakala, Tript 2, and Lijinsky can help to further resolve this disagreement. Although not shown in Figure 1, among 14 chemicals classified as non-carcinogenic by both Tennant and Weisburger, two proved to be carcinogenic in the rodent bioassays.

Overall, hybridizing multiple predictions via the tree in Figure 1 leads to a slightly smaller average error rate (22.7% resubstitution error) than the best individual (Tennant) system alone has. Although tree model cross-validation with generalized degrees of freedom (Ye 1998), could yield better estimates of the true error rate for the tree by better correcting for bias due to over-fitting, the simple resubstitution error estimates suffice to illustrate the following key points:

e *Nonmonotonicity:* The best combination of expert system predictions need not be monotonic, i.e., the probability that a chemical is a rodent carcinogen may be decreased by learning that a system classifies it as such. For example, the Lijinsky node at the bottom of the tree shows a significant inverse relation between predicted and true classes, with the conditional probability that a chemical is a rodent carcinogen falling from 63.6% to 50%, given the results of the preceding tests in the tree, when the Lijinsky system predicts that it is a carcinogen. Although the sample sizes are small, the same phenomenon can occur in trees with more cases. The reason is that prediction errors made by different tests can interact in strong, potentially counter-intuitive ways. Exploiting such interactions enables classification tree hybrids to make improved predictions. This is a novel way to combine predictions from different sources, however, and it can violate many of the principles (such as unanimity and monotonicity) often proposed as normative axioms in previous approaches to combining predictions from different expert sources (Clemen 1989.)

*High-order interactions among tests.* The best predictions that can be achieved from a set of tests such as those in Table 1 may depend on interactions of many individual tests (e.g., five in Figure 1). Thus, the information contained in the one or two individually "best" tests does not subsume the useful information in the other tests.

o *Information value synergies:* The information value of a particular test can depend strongly on what other tests have been performed. For example, even though the predictive value of the Lijinsky test is only slightly better than random guessing when it is considered in isolation (i.e., its misclassification rate is close to 50%), it can help to identify high-probability carcinogens when used in the context of other tests in Figure 1.

F *No dominance:* It may be natural to think of some predictive systems as being strictly "better than" or "more informative than" others, so that when predictions from the best systems are known, those from less good systems should be ignored. Figure 1 suggests that relations among predictions can be more complicated, with even relatively weak predictive systems being able to add value for some combinations of predictions by the better systems. A formal basis for comparing predictive systems using classification trees is introduced in the next section.

That a combination of predictions from diverse sources can out-perform any of the individual sources is perhaps to be expected, based on much previous management science research on optimally combining or aggregating expert predictions (Clemen 1989). However, a novel feature of our approach is the use of classification trees, rather than analytic aggregation or averaging formulas, to combine the predictions from different AI and statistical prediction systems. This allows higher-order interactions among the prediction errors from different systems to be exploited in constructing combined predictions. As suggested by Figure 1, such interactions are potentially valuable: common combination methods based only on the variances and covariances of predictions from different sources may leave potentially valuable information unused. The tree approach also brings within a natural probabilistic framework systems that do not by themselves yield probabilistic predictions. This is done by treating their deterministic predictions (typically "carcinogen", "non-carcinogen" or "unable to make a determination") as values on which the combined, probabilistic prediction is conditioned.

## 3 METHODS AND DATA

Over the past three decades, a variety of increasingly sophisticated artificial intelligence and statistics methods have been brought to bear on the problem of predicting chemical carcinogens. Several well-developed approaches were recently evaluated in blind tests (Benigni 1996, Bristol 1996), i.e., they were used to predict whether various chemicals would be found to be rodent carcinogens in ongoing (typically, two-year) bioassay experiments in mice and rats. The predictions were published before the results of the experiments were known.

Key approaches tested for predictive accuracy include the following:

*Structure-activity relation (SAR) programs.* These consider physical and electronic properties, three-dimensional molecular structure, and molecular topological indices, indicating key invariants such as graph-theoretic structures (Perrotta 96) associated with DNA reactivity. The SAR and quantitative SAR (QSAR) approaches taken to date include:

q Benigni's method combines the structural alerts of Tennant and Ashby (below) with Bakale's coefficient of electrophilic reactivity, denoted Ke, (below) to obtain a QSAR score.

o CASE / MULTICASE (Cunningham 1998) is a Bayesian QSAR statistical expert system that uses statistically selected relations among attributes of chemical structures to identify substructures useful for predicting probable carcinogenicity. It differs from earlier QSAR expert systems in that it fully automates the selection of chemical substructures to be considered, rather than requiring a human user to select them from a library.

s COMPACT (Lewis 1998) is a QSAR system that calculates approximate molecular dimensions and molecular and electronic structures via "Computer-optimized molecular parametric analysis for chemical toxicity" to predict whether a chemical will be metabolically activated to a carcinogenic chemical by specific enzymes.

s DEREK (Marchant 1996) is a rule-based expert SAR system based on "deductive estimation of risk from existing knowledge" obtained from expert chemists.

n TOPKAT (Enslein 1990) applies statistical regression and discriminant analysis to chemical structural attributes to obtain SAR rules.

a Bakale Ke (Bakale 1992). This uses a single measured parameter, Ke = chemical electrophilic reactivity, to predict carcinogenic potential.

p Weisburger (an unpublished SAR system encoding expert intuition)

*Activity-activity relation (AAR) programs.* These use the spectrum of biological responses in relatively inexpensive assays (e.g., bacteria mutation tests or short-term toxicity and tests) to predict biological activities in more expensive and relevant systems (e.g., two-year rodent cancer bioassays). Some AAR systems, including that of Tennant and Ashby, also use any available data from previous cancer bioassays.

AAR systems evaluated for predictive validity (Benigni 1996, Bristol 1996) include:

) Tennant and Ashby's AAR system (Tennant 1990). This uses correlations among attributes of chemical structures, short-term mutagenicity test results (e.g., in *Salmonella*), rodent subchronic toxicity outcomes, and carcinogenicity test results if available. It identifies "structural alerts" indicating possible carcinogenicity. This system has been fine-tuned by its expert authors based on reviews of biological response profiles and chemical structures for over 300 chemicals. It incorporates much of their intuition. The system requires a human expert, i.e., it is not fully automatic.

i RASH (Jones 1996). The "rapid screening of hazards" method predicts carcinogenic potential based on the observed relative potencies of tested chemicals in different short-term bioassays. It is not fully automatic, but instead requires a human expert to select relevant comparisons.

t TRIPT (Bahler 1993) performs "tree and rule induction for predictive toxicology" via the machine learning algorithm C4.5, applied to the factors considered in the Tennant system.

c PROGOL (King 1996) applies inductive logic programming (ILP) to relational descriptions of chemical structures to induce simple, interpretable rules for SAR structural alerts.

Other methods for which predictions have been recorded include Fuzzy Adaptive Least Squares (Moriguchi 1996) and the unpublished predictive systems of Lijinsky and Weisburger.

Table 1 summarizes data on the outcomes of the different prediction methods (Bristol 1996) so that other AI and statistics researchers can try their own programs on the chemical carcinogenicity prediction task. It consists of the predictions for 44 chemicals made by different prediction systems using the above methods, ranging from statistical (e.g., TOPKAT) to rule-based expert systems to machine-learning (e.g., TRIPT)

approaches. Table 1 also presents the actual carcinogenicity outcomes observed for each chemical in both sexes of both mice and rats, based on experiments completed after the predictions were made. The codes for bioassay outcomes in individual species (M = mouse, R = rat) and sexes (M = male, F = female) are: CE = clear evidence of carcinogenicity; SE = some evidence; EE = equivocal evidence; NE = negative evidence

We analyzed these data, using the main classification tree algorithms implemented in KnowledgeSeeker™ with automatic Bonferroni adjustments to protect against multiple testing bias (Biggs 1991), to construct several trees that yield improved predictions. First, at the risk of over-training on the sample data, we conducted exploratory analyses of the whole data set (using KnowledgeSeeker's "Exhaustive" tree-growing algorithm) to detect patterns in errors across the different prediction methods. Then we used various random partitions of the 44 chemicals into training and test sets to assess the performance of the tree hybridization approach. (Typically, we used 29 chemicals to train and 15 to test.) A best-informed (S*) tree, as defined in the following section, was used to generate "hybridized" predictions from predictions already made by the different methods, along with statistics on their errors in the test set.

## 4 RESULTS AND DISCUSSION

Two major practical goals of new efforts in this field are to reduce the costs and increase the accuracy of predictive classification of chemicals. The costs are driven largely by *in vivo* testing implying that SAR and QSAR methods tend to be much less expensive than AAR methods, especially when the latter involve results of lengthy *in vivo* toxicity tests. By assembling batteries of tests that place relatively inexpensive tests first, the expected costs of reaching a classification decision with a specified level of confidence can sometimes be dramatically reduced. Indeed, this principle has been used in recent algorithms and heuristics for minimizing average costs of testing (Cox 1994). On the other hand, incorporating a few very expensive tests, such as a long-term cancer bioassay for a single rodent species and sex, can lead to dramatic improvements in accuracy if the predictive tests are used to select the test species and sex and to help interpret the results. The following paragraphs present our main findings on how classification trees can be used to improve cost-accuracy trade-offs and to compare different prediction systems.

### 4.1 REDUCING CLASSIFICATION COSTS

Figure 2 shows a tree with a resubstitution error rate of only 4.5%, far less than the 22.5% achievable if no long-term animal cancer bioassays are used (see Figure 1). It illustrates the value of combining the predictions from several systems with the results of a single long-term bioassay (either MM = male mice or FR = female rats), where the Tennant system's prediction is used to select which animal bioassay to perform. After performing the tests indicated in this tree, additional bioassays in other sexes or species do not improve predictive accuracy further.

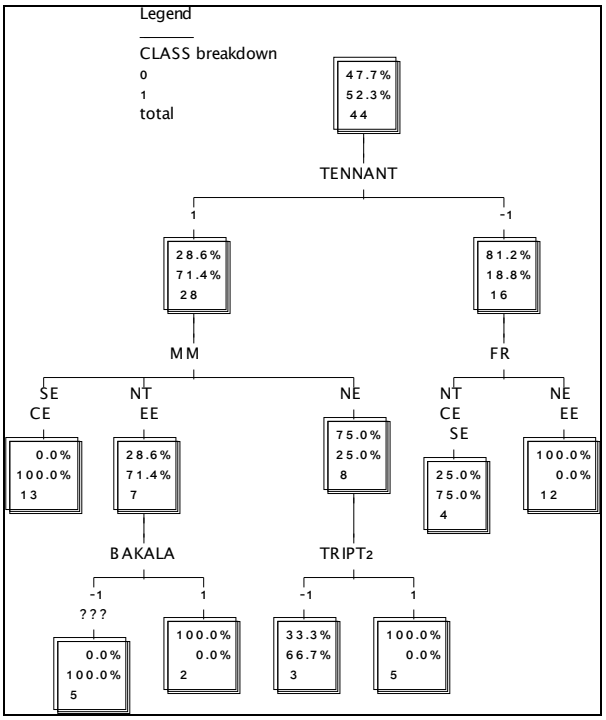

Figure 2: Predictions Help to Economize on Bioassays

Thus, classification tree analysis reveals that a chemical's carcinogenicity class is *conditionally independent* of the remaining information (predictions from other systems and bioassay experiment outcomes), at least as far as the tree-growing algorithm can discover, once one of the seven leaf nodes in Figure 2 has been reached.

It is noteworthy that the resubstitution error rate from the male mouse (MM) bioassay alone is 20%, while from the female rate (FR) test alone it is 30%. From the two together, it is 14%. Yet, hybridizing these tests with the other imperfect predictions in Figure 2 (Tennant, Bakala, and Tript2), which taken together have a joint error rate of 25%, produces a hybrid classification scheme with an error rate of less than 5%. Interpretively, this suggests that the MM and FR tests provide information that is approximately orthogonal to (complementary to) the information provided by the set of tests {Tennant, Bakala, Tript2}. This interpretation is strengthened by observing that, after constructing a tree using these three variables, the MM and FR tests will still enter at the bottom of that tree if allowed to, implying that a chemical's class is *not* conditionally independent of MM and FR, given the predictions from

{Tennant, Bakala, Tript2}. Once MM and FR have entered, however, no other variables in Table 1will, showing that the set {Tennant, Bakala, Tript2, MM, FR} is *sufficient* for the full set of variables. Searching the set of possible trees shows that this is a *minimal sufficient set* (i.e., none of its subsets has the sufficiency property) and that no other set achieves both a smaller error rate and a smaller cost (assuming that animal bioassays cost more than AAR methods and that AAR methods cost more than SAR or QSAR methods.) See (Cox 1994) for search algorithms and heuristics for finding cost-effective trees.

## 4.2 COMPARING PREDICTIVE SYSTEMS

The classification tree framework for combining predictions contributes a new technique for comparing information sources. Roughly speaking, one source of predictions may be considered "better-informed" or "more valuable" than another if *every* rational decision-maker would prefer to obtain an observation from the first instead of the second before making a decision. (This assumes that the payoff or utility from the decision depends on the true state, e.g., carcinogenic or not, for the chemical being classified.) This comparative binary relation generally yields a partial ordering of information sources when the characteristics of the sources (i.e., probabilities of outputs given the true states) are known. While several equivalent characterizations have been given for determining when one information source is more valuable than another, the classification tree framework provides a simple test that does not require a priori knowledge of the probability characteristics of the sources. Call source S1 *at least as well informed as* source S2 if the classification of a chemical is conditionally independent (CI) of the prediction from S2, given the prediction from S1. S1 is *better informed* than S2 if S1 is at least as well informed as S2 but not *vice versa.* The classification tree test for this relation is as follows.

| Comparing Predictive Systems via a Tree Algorithm | |
|---|---|
| C | Take the correct classification of a chemical (e.g., carcinogen or not) as the dependent variable. |
| c | Split the correct classification on the output (predictions) from S1 to form a one-split tree. Call this tree T1. |
| t | Allow tree T1 to be extended by splitting each of its leaves on the predictions from S2. |
| l | If S2 does not enter the tree when this is done, but if splitting the chemical classification first on S2 and then on S1 results in S1 entering the tree below S2, then S1 is better informed than S2. Interpretively, S1 is "sufficient for" S2, but not *vice versa;* see DeGroot, 1970, Chapter 14. |

This characterization of comparative expertise defines a partial ordering that complements earlier ones in the statistical decision theory literature (DeGroot, 1970, 433-439). It can readily be extended to compare *subsets* of variables (or tests, or predictive systems), by treating each subset in turn as the allowed set of independent variables entering a classification tree analysis. For example, after conditioning (by growing a tree, T1) on {Weisburger, RASH, COMPACT}, Tennant and Tript2 will still enter at the bottom of the tree if allowed to. Conversely, after conditioning on {Tennant, Bakala, Tript2}, Weisburger, RASH, and COMPACT will all still enter if allowed to. Thus, neither set is more informative than the other: they are complementary.

## 4.3 OPTIMALLY COMBINING PREDICTIONS

The ability to compare prediction systems based on trees suggests constructing a new, *best-informed source of* predictions, say, S*, by combining the predictions from individual sources into a tree such no other tree composed from these sources is better-informed than S*. A useful approximate construction heuristic that uses standard classification tree algorithms is the following. Grow an initial tree myopically, e.g., by choosing the strongest predictors of correct classes in the training set first. Then refine it by making pairwise swaps of variables below the root node with the root node variable until no further improvements (defined as reduction of average prediction error rate in the test set) can be found. Apply this tree-improvement routine to several initial trees formed by random selection of the top few candidate splits at each node. The result is often a tree that (a) yields the smallest achievable average prediction error in the test set; and (b) does so using an efficient set of variables, i.e., a set of variables such that any proper subset yields significantly deteriorated performance. This tree is a practical estimate or approximation to S*, the best-informed predictor of chemical class that can be formed from the sources considered.

In our experience, as discussed in Section 2, the best-informed source if often inconsistent with axioms that have sometimes been proposed in the management science literature for combining expert predictions (e.g., the "unanimity" axiom, according to which S* should make the same prediction as its component sources when all of them agree). However, it is easy to demonstrate by examples that, in these cases, the axioms are not useful, whereas S* is, in making the best possible predictions. Thus, it appears that classification trees may offer a useful general alternative to previous methods of combining expert predictions, as well as making more specific contributions to chemical carcinogen prediction.

### 4.4 OTHER FINDINGS

Other findings from our classification tree analysis of Table 1 include the following.

l The different methods are much better predictors of each other than of the true classification of the chemicals. This suggests that there is important information captured in the rodent cancer bioassays that is not captured in the predictive methods currently in use. As shown in Figure 2, such complementary information can be exploited to reduce the number of expensive bioassays performed.

e Combining predictions from different predictive methods leads to a best-informed tree that improves on the predictive accuracy of any single method. The best-informed tree is not unique, however.

b When only the least expensive (SAR or QSAR, but not AAR) tests are considered, the best hybrid tree classifier based on {Bakala's Ke, MULTI-CASE, COMPACT, DEREK, TOPKAT} has a 20% resubstitution misclassification error rate.

r A key goal of carcinogen prediction has been to identify a battery of low-cost tests and assays that would collectively be as informative as the much more expensive rodent bioassays about the likely carcinogenicity of chemicals. Classification trees were applied to test how well this goal has been achieved. Whether carcinogenicity in each species and sex is conditionally independent of carcinogenicity in the other three, given the results of all predictive methods, can be tested by the two-phase tree-growing procedure outlined above. It turns out that the other carcinogenicity bioassays contain relevant information not captured in the predictive methods (i.e.., the predictive methods being used are not sufficient for the carcinogenicity experiments.)

n On the positive side, classification trees show that carcinogenicity of a chemical in a specific rodent species and sex can be predicted as well from carcinogenicity testing results in one other species and sex and a few (typically two) of the prediction methods as it can be from results of carcinogenicity testing in all three other species-sex combinations

t The classification tree method suggests the possibility (and provides a constructive algorithm) for combining whole-animal carcinogenicity testing with less expensive predictive methods to obtain predictions of human carcinogenicity that are at least as informed as methods based on more extensive whole-animal testing in additional species and sexes.

## 5 SUMMARY AND CONCLUSIONS

In summary, we have identified a tree-based approach to combining the results of multiple tests to reduce test costs (e.g., by using results of less expensive tests to determine which expensive ones to perform) and to reduce error rates by hybridizing predictions based on complementary information. The approach appears promising for the data in Table 1. It can be implemented using standard classification tree software. However, some important open issues remain. These include the following.

1. *Tracking concept drift*. The classification tree analysis revealed that year of completion of peer review of rodent cancer bioassays is itself quite informative about the likelihood that a chemical is a rodent carcinogen. Roughly 30% of chemicals reviewed in 1990, 60% of those reviewed in 1991-1993, and 100% of the (three) chemicals reviewed in 1994 and 1995 were found to be rodent carcinogens. Thus, the proportion of rodent carcinogens among chemicals selected for long-term cancer bioassays may be increasing over time. (Indeed, if year of review is used as the sole predictor of rodent carcinogenicity, the sample misclassification error rate is 36%, lower than for several of the predictive systems.) When training and test sets are obtained by partitioning chemicals according to the year in which a peer-reviewed cancer bioassay was completed, it appears that the first chemicals tested (mainly genotoxic ones) yield trees that are especially weak predictors of the carcinogenicity of later chemicals (which contain more non-genotoxic ones). Thus, when the concept being learned "drifts" over time (e.g., away from genotoxic and toward non-genotoxic carcinogens, in this case), it is important to make sure that the training set is balanced (or re-balanced) to adequately emphasize the components that are to be predicted.

2. *Formal cost-optimization.* This paper has emphasized construction of best-informed sources from several less-informed sources. As briefly mentioned, a useful extension would be to assign costs to the different tests and seek a *minimum-expected cost tree* that balances the costs of testing against the costs of decision error. Computational complexity results and practical heuristics are available for such problems (Cox 1994).

3. *Latent variables.* A potentially desirable approach to predicting chemical carcinogenicity is to allow for hierarchical concept-learning, including induction of latent variables (such as "genotoxic carcinogen"). Such variables do not arise as

Boolean combinations of attribute values, but may greatly simplify the interpretation of attribute value combinations. It may be worthwhile to extend classification tree algorithms to partition training sets into relevant and irrelevant exemplars, based on hypothesizing a latent variable (e.g., the "genotoxic" classification) that is related to the observed attribute values but not directly measured. For example, we have found that classification trees can provide powerful predictors of mineral oil carcinogenicity, with clear advantages compared to older statistical methods, if latent variables can first be used to partition the training and test sets into relatively homogeneous subsets.

**Acknowledgment**

This research was stimulated by and has benefited from many discussions with Dr. Michael Bird of Exxon Biomedical Science, Inc. (EBSI). I am grateful to Dr. Bird and to EBSI for encouraging and supporting my research on better ways to use information about biological response profiles to predict the likely health effects of chemicals.

**References**

Ashby, J., and D. Paton, 1993. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogens. *Mutation Research,* 286, 3-74.

Bahler D. and D.W. Bristol, 1993. The induction of rules for predicting chemical carcinogenesis in rodents. *Ismb*; **1**:29-37

Bakale G. and R.D. McCreary R.D., 1992. Response of the ke test to NCI/NTP-screened chemicals. II. Genotoxic carcinogens and non-genotoxic non-carcinogens. *Carcinogenesis*; **13** (8): 1437-45

Benigni, R., 1997. The first US National Toxicology Program exercise on the prediction of rodent carcinogenicity: definitive results. *Mutation Research*; **387**(1):35-45

Benigni, R., 1996. Predicting chemical carcinogenesis in rodents: The state of the art in light of a compara-tive exercise. *Mutation Research,* 334, 103-113.

Benigni, R., 1991. QSAR prediction of rodent carcinogenicity for a set of chemicals currently bioassayed by the US National Toxicology Program.
*Mutagenesis* 1991 Sep;6(5):423-5

Biggs, D., B. de Ville, E. Suen, 1991. A method of choosing multiway partitions for classification and decision trees. *J. Applied Statistics,* 18, 1, 49-62.

Bristol D.W., J.T. Wachsman, A. Greenwell, 1996. The NIEHS Predictive-Toxicology Evaluation Project. *Environ. Health Perspectives;* **104** Suppl 5:1001-10.

Chevalier, S., R.A. Roberts RA, 1998. Perturbation of rodent hepatocyte growth control by nongenotoxic hepato-carcinogens: mechanisms and lack of relevance for human health. *Oncology Reports*; 5(6):1319-27.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography," (with discussion). *International Journal of Forecasting,* **5**, 559-583.

Cox, L.A., Jr., 1997. Does diesel exhaust cause human lung cancer? *Risk Analysis*, **17**, 6, 807-829.

Cox, L.A., Jr., and Y. Qiu, 1994. "Minimizing the expected costs of classifying patterns by sequential costly inspections," in Cheeseman, P. and R.W. Olford (Eds), *Selecting Models from Data.* Springer-Verlag, *Lecture Notes in Statistics,* Volume 89, pp. 339-350. New York.

Cunningham AR, G. Klopman, H.S. Rosenkranz, 1998. Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutation Research* **31**; 405(1):9-27.

DeGroot, M.H., 1970. *Optimal Statistical Decisions*. McGraw-Hill.

Enslein K., B.W. Blake, H.H. Borgstedt, 1990. Prediction of probability of carcinogenicity for a set of ongoing NTP bioassays. *Mutagenesis*; **5**(4): 305-6

Gold, L.S., T.H. Slone, N.B. Manley, L. Bernstein, 1991. Target organs in chronic bioassays of 533 chemical carcinogens. *Environmental Health Perspectives*, **93**, 233-246.

Jones T.D. and C.E. Easterly, 1996. A RASH Analysis of National Toxicity Program Data: Predictions for 30 Compounds to Be Tested in Rodent Carcinogenesis Experiments. *Environmental Health Perspectives*; **104** S (5): 1017-30

King, R.D., and A. Srinivasan, 1996. Prediction of rodent carcinogenicity bioassays from molecular structure using inductive logic programming. *Environmental Health Perspectives,* **104**S(5): 1031-1039.

Lewis D.F., C. Ioannides, D.V. Parke, 1998. Further evaluation of COMPACT, the molecular orbital approach for the prospective safety evaluation of chemicals. *Mutation Research*,**13**;412(1):41-54

Marchant, C.A., 1996. Prediction of rodent carcinogenicity using the DEREK System for 30 chemicals currently being tested by the National Toxicology Program. *Environmental Health Perspectives;* **104**S(5):1065-73

Moriguchi I, H. Hirano H, 1996. Prediction of the rodent carcinogenicity of organic compounds from their chemical structures using the FALS method. *Environmental Health Perspectives*; **104S**(5):1051-8

Perrotta A., D. Malacarne, M. Taningher, R. Pesenti, M. Paolucci, S. Parodi, 1996. A computerized connectivity approach for analyzing the structural basis of muta-genicity in Salmonella and its relationship with rodent carcinogenicity. *Environ Mol Mutagen*; **28**(1): 31-50

Taningher, M. *et al.,* 1997. Methods for predicting carcinogenic hazards: New opportunities coming from recent developments in molecular oncology and SAR studies. *Mutation Research,* 391, 3-32.

Templin, M.V., A.A. Constan, D.C. Wolf, B.A. Wong, B.E. Butterworth, 1998. Patterns of chloroform-induced regenerative cell proliferation in BDF1 mice correlate with organ specificity and dose-response of tumor formation. *Carcinogenesis;* **19**(1):187-93.

Williams G.M., M.J. Iatropoulos, J.H. Weisburger JH, 1996. Chemical carcinogen mechanisms of action and implications for testing methodology. *Experimental Toxicology and Pathology;* **48**(2-3):101-11.

Ye, J., 1998. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association,* **93**(441), 120–131 http://www.amstat.org/publications/jasa/abstracts_98/YE.HTM

Yoshikawa, K., 1996. Anomalous nonidentity between Salmonella genotoxicants and rodent carcinogens: nongenotoxic carcinogens and genotoxic non-carcinogens. *Environmental Health Perspectives;* **104** (1):40-6.

**References with Web Links**


Ashby, J., and D. Paton, 1993. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogens. *Mutation Research,* 286, 3-74. http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=7678908&form=6&db=m&Dopt=b

Benigni, R., 1995. Predicting chemical carcinogenesis in rodents: The state of the art in light of a comparative exercise. *Mutation Research,* 334, 103-113.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=7528333&form=6&db=m&Dopt=b

Bakale G. and R.D. McCreary R.D., 1992. Response of the ke test to NCI/NTP-screened chemicals. II. Genotoxic carcinogens and non-genotoxic non-carcinogens. *Carcinogenesis*;**13**(8):1437-45
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=1499095&form=6&db=m&Dopt=b

Biggs, D., B. de Ville, E. Suen, 1991. A method of choosing multiway partitions for classification and decision trees. *J. Applied Statistics,* 18, 1, 49-62.

Bristol D.W., J.T. Wachsman, A. Greenwell, 1996. The NIEHS Predictive-Toxicology Evaluation Project. *Environ. Health Perspectives;***104** Suppl 5:1001-10. http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933048&form=6&db=m&Dopt=b

Chevalier S., R.A. Roberts RA, 1998. Perturbation of rodent hepatocyte growth control by nongenotoxic hepato-carcinogens: mechanisms and lack of relevance for human health. *Oncology Reports*; 5(6):1319-27.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9769363&form=6&db=m&Dopt=b

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography," (with discussion). *International Journal of Forecasting,* **5**, 559-583.
http://www.duke.edu/~clemen/pubs.htm

Cox, L.A., Jr., 1997. Does diesel exhaust cause human lung cancer? *Risk Analysis*, **17**, 6, 807-829.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9463933&form=6&db=m&Dopt=b

Cox, L.A., Jr., and Y. Qiu, "Minimizing the expected costs of classifying patterns by sequential costly inspections," in Cheeseman, P. and R.W. Olford (Eds), *Selecting Models from Data.* Springer-Verlag, *Lecture Notes in Statistics,* Volume 89, pp. 339-350. New York, 1994.

Cunningham AR, G. Klopman, H.S. Rosenkranz, 1998. Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutation Research* **31**; 405(1):9-27.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9729240&form=6&db=m&Dopt=b

DeGroot, M.H., 1970. *Optimal Statistical Decisions*. McGraw-Hill.

Gold, L.S., T.H. Slone, N.B. Manley, L. Bernstein, 1991. Target organs in chronic bioassays of 533 chemical carcinogens. *Environmental Health Perspectives*, **93**, 233-246. http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=1773795&form=6&db=m&Dopt=b

King, R.D., and A. Srinivasan, 1996. Prediction of rodent carcinogenicity bioassays from molecular structure using inductive logic programming. *Environmental Health Perspectives,***104**S(5):1031-1039.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933051&form=6&db=m&Dopt=b

Lewis D.F., C. Ioannides, D.V. Parke, 1998. Further evaluation of COMPACT, the molecular orbital approach for the prospective safety evaluation of chemicals. *Mutation Research*,**13**;412(1):41-54
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9508363&form=6&db=m&Dopt=b

Marchant, C.A., 1996. Prediction of Rodent Carcinogenicity Using the DEREK System for 30 Chemicals Currently Being Tested by the National Toxicology Program. *Environmental Health Perspectives;* **104**S(5): 1065-73
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933056&form=6&db=m&Dopt=b

Moriguchi I, H. Hirano H, 1996. Prediction of the rodent carcinogenicity of organic compounds from their chemical structures using the FALS method. *Environmental Health Perspectives***; 104S**(5):1051-8

Perrotta A., D. Malacarne, M. Taningher, R. Pesenti, M. Paolucci, S. Parodi, 1996. A computerized connectivity approach for analyzing the structural basis of muta-genicity in Salmonella and its relationship with rodent carcinogenicity. *Environ Mol Mutagen*; **28**(1):31-50
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8698045&form=6&db=m&Dopt=b

Taningher, M. *et al.,* 1997. Methods for predicting carcinogenic hazards: New opportunities coming from recent developments in molecular oncology and SAR studies. *Mutation Research,* 391, 3-32.

Templin, M.V., A.A. Constan, D.C. Wolf, B.A. Wong, B.E. Butterworth, 1998. Patterns of chloroform-induced regenerative cell proliferation in BDF1 mice correlate with organ specificity and dose-response of tumor formation. *Carcinogenesis;* **19**(1):187-93.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9472711&form=6&db=m&Dopt=b

Williams G.M., M.J. Iatropoulos, J.H. Weisburger JH, 1996. Chemical carcinogen mechanisms of action and implications for testing methodology. *Experimental Toxicology and Pathology;* **48**(2-3):101-11.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8672863&form=6&db=m&Dopt=b

# SELECTED ABSTRACTS

ASHBY 93
The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures.
Ashby J, Paton D
ICI Central Toxicological Laboratory, Macclesfield, Ches., UK.

Gold and her colleagues have tabulated the results of rodent bioassays on 522 chemicals and have analysed the data. The present study complements those analyses by providing a perspective from the viewpoint of the chemical structure of the carcinogens. The chemical structure of each of the carcinogens is displayed and the Gold database is represented with the test agents as the primary variable. **The carcinogens are gathered into six chemical classes and each chemical is assessed for structural alerts to DNA reactivity**. The database is then analysed using an integration of the following parameters: bioassay in rat, mouse or both; structural alert status; chemical class; sites and multiplicity of carcinogenesis, and trans-species carcinogenicity. A series of Figures is presented that enables rapid acquaintance with what represents the core database of rodent carcinogenicity. The several analyses presented combine in endorsing the reality of two broad classes of rodent carcinogen--presumed DNA-reactive and others (putative genotoxic and non-genotoxic carcinogens, but semantics have been largely avoided). Vainio and his colleagues have tabulated 55 situations in which humans have succumbed to chemically induced cancer, and have listed the tissues affected. This database of human carcinogens has been analysed in the present study as done for the rodent carcinogen database, and comparisons made between the two. The predominance of putative genotoxic carcinogens in the human database was confirmed, as was the reality of putative non-genotoxic carcinogenicity in humans. It is concluded that putative genotoxic rodent carcinogenesis can be correlated both with chemical structure and the extent and nature of the induced effect, and that it is of clear relevance to humans. **In contrast, it is concluded that putative non-genotoxic rodent carcinogenesis is more closely related to the test species than to the test chemical, and that it is essentially unpredictable in the absence of mechanistic models. In the absence of such models nongenotoxic carcinogenic effects should be extrapolated to humans with caution.** Progress in the accurate prediction and extrapolation of rodent carcinogenicity will be helped by a common, if only temporary, enabling acceptance that not all carcinogens are intrinsically genotoxic.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=7678908&form=6&db=m&Dopt=b

BAHLER 93
Bahler D. and D.W. Bristol, 1993. The induction of rules for predicting chemical carcinogenesis in rodents. *Ismb*;**1**:29-37
Dept. of Computer Science, North Carolina State University, Raleigh 27695-8206, USA.

This paper presents results from an ongoing effort in applying a variety of induction-based methods to the problem of predicting the biological activity of noncongeneric (structurally dissimilar) chemicals. It describes initial experiments, the long-term goal of which is to assist toxicologists, cancer researchers, regulators, and others to predict the toxic effects of chemical compounds. We describe a series of experiments in **tree and rule induction from a set of example chemicals whose carcinogenicity has been determined from long-term animal studies**, and compare the resulting classification accuracy with eight published human and computer predictions for a common set of 44 test chemicals. **The accuracy of our system is comparable to the most accurate human expert prediction yet published, and exceeds that of any of the computer-based predictions in the literature.** The induced rules provide confirmation of current expert heuristic knowledge in this domain. These early results show that an inductive approach has excellent potential in predictive toxicology.

PMID: 7584348, UI: 96038948
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=7584348&form=6&db=m&Dopt=b

BAKALE 92
Bakale G. and R.D. McCreary R.D., 1992. Response of the ke test to NCI/NTP-screened chemicals. II. Genotoxic carcinogens and non-genotoxic non-carcinogens. *Carcinogenesis*;**13**(8):1437-45

Case Western Reserve University, Department of Radiology, Cleveland, OH 44106-5000.

A physico-chemical carcinogen-screening test was used to measure the rate constants of electron attachment, kes, of 105 chemicals that had been screened in long-term rodent bioassays and short-term in vitro tests by the NCI/NTP. In the ke test, a pulse-conductivity technique is used to generate and monitor the decay of excess electrons that serve as nucleophilic surrogates for the target tissue of rodents. Of the 61 chemicals that had been found to be rodent carcinogens as well as Salmonella mutagens, 36 yield kes that are equal to or greater than the diffusion-controlled ke of carbon tetrachloride and are considered to be positive ke test responses. In contrast, 29 of the remaining 44 chemicals that are putative non-carcinogens and non-mutagens yield kes that are negative ke test responses. These results are combined with the ke responses of 46 non-mutagenic carcinogens and 20 mutagenic non-carcinogens that were reported earlier and are evaluated to determine the degree to which the measure of electron-accepting capacity that ke provides complements or overlaps the electrophilicity or DNA reactivity of chemicals that is indicated by positive mutagenicity responses in the Ames Salmonella tester strains or by positive structural alerts, S/As, of the chemicals. **The combined ke test results indicate that the overall predictivity of the ke test is comparable to and complements the Ames Salmonella test and S/As in identifying rodent carcinogens.** Moreover, the electrons serve as non-discriminate nucleophilic targets for both genotoxic and non-genotoxic electron-accepting molecules and appear to attach with equal efficiency to carcinogens that are active in various tissues of rodents. This property of excess **electrons suggests that the predictivity of the ke test could be enhanced by combining the measured ke with an appropriate lipophilicity or pharmacokinetic parameter**. A pre-chemical electron-transfer step that had been proposed to precede chemical interactions between the carcinogen and target tissue is discussed in light of recent developments in electron-donor/-acceptor chemistry and in the application of structure--activity relationships to identify carcinogens.

PMID: 1499095, UI: 92361970
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=1499095&form=6&db=m&Dopt=b

BENIGNI 95
Mutat Res 1995 Feb;334(1):103-13
Predicting chemical carcinogenesis in rodents: the state of the art in light of a comparative exercise.
Benigni R
Istituto Superiore di Sanita, Laboratory of Comparative Toxicology and Ecotoxicology, Rome, Italy.

Within a recent comparative exercise, different approaches to the prediction of rodent carcinogenicity were challenged on a common set of chemicals bioassayed by the U.S. National Toxicology Program. The approaches were of very different natures. Some prediction systems looked for relationships between carcinogenicity and other, more quickly detectable biological events (activity-activity relationships, AAR). Some approaches tended to find structure-activity relationships (SAR). To give an objective evaluation of the results of the exercise, we have analyzed the rodent results and the predictions with the multivariate data analysis methods. **The calculated performances varied according to the adopted carcinogenicity classification of the chemicals**. When the four rodent results were summarized into a final + or - call, the Tennant approach (AAR method) showed the best performance (about 75% accuracy), whereas the best SAR systems had 60-65% accuracy. **A common limitation of almost all the systems was the lack of specificity (too many false positives).** Based on these results**, better concordance was obtained when the input information was the very costly (and closer to the final endpoint) biological data**, rather than the inexpensive (and farther from the endpoint) knowledge of the chemical structure. However, when the rodent results were summarized into a carcinogenicity classification that maintained, to some extent, the gradation intrinsic to the original experimental data, the performance of the AAR systems declined, and the SAR approaches showed a better performance. The difficulty in evaluating the various approaches was further complicated because of a fundamental difference in the approaches themselves: some approaches were 'pure' prediction methods (i.e. their predictions were rigorously based on information not inclusive of carcinogenicity); other approaches (e.g. Tennant, Weisburger) used 'mixed' information, inclusive of known carcinogenicity results from experiments performed before the NTP bioassays. As far as the SAR systems are concerned, their sets of predictions showed a fundamental similarity. This happened in spite of the extremely different procedures adopted to treat the chemical formula (initial information): very simple calculations (Benigni), intuition of the experts (Weisburger and Lijinsky), sophisticated computer programs (TOPKAT and CASE). The **results of the Bakale Ke method, based on the experimental measurement of the**

**chemical electrophilicity, and of the Salmonella typhimurium mutagenicity assay were similar to the patterns of predictions of the SAR methods**.

PMID: 7528333, UI: 95098052
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=7528333&form=6&db=m&Dopt=b

BENFENATI 96
Toxicology 1997 May 16;119(3):213-25
Computational predictive programs (expert systems) in toxicology.
Benfenati E, Gini G

Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy. benfenati@irfmn.mnegri.it

The increasing number of pollutants in the environment raises the problem of the toxicological risk evaluation of these chemicals. Several so called expert systems (ES) have been claimed to be able to predict toxicity of certain chemical structures. Different approaches are currently used for these ES, based on explicit rules derived from the **knowledge of human experts** that compiled lists of toxic moieties for instance in the case of programs called HazardExpert and **DEREK** or relying on **statistical approaches, as in the CASE and TOPKAT programs.** Here we describe and compare these and other intelligent computer programs because of their utility in obtaining at least a first rough indication of the potential toxic activity of chemicals.

BENIGNI 91
*Mutagenesis* 1991 Sep;6(5):423-5
QSAR prediction of rodent carcinogenicity for a set of chemicals currently bioassayed by the US National Toxicology Program.
Benigni R
Laboratory of Comparative Toxicology and Ecotoxicology, Istituto Superiore di Sanita, Rome, Italy.

A **QSAR model based on the combination of two molecular descriptors--estimated electrophilic reactivity and Ashby's structural alerts**--was used to predict the carcinogenicity of 44 chemicals currently bioassayed by the US National Toxicology Program. These predictions will be compared with the rodent carcinogenicity assay results as the assays are completed.
PMID: 1795649, UI: 92178057
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=1795649&form=6&db=m&Dopt=b

BENIGNI 96
Environ Health Perspect 1996 Oct;104S(5):1041-4
Prediction of Rodent Carcinogenicity of Further 30 Chemicals Bioassayed by the U.S. National Toxicology Program.
Benigni R, Andreoli C, Zito R
Laboratory of Comparative Toxicology and Ecotoxicology, Istituto Superiore di Sanita', Rome, Italy

Recently the U.S. National Toxicology Program (NTP) sponsored a comparative exercise in which different prediction approaches (both biologically and chemically based) were challenged for their predictive abilities of rodent carcinogenicity of a common set of chemicals. The exercise enjoyed remarkable scientific success and stimulated NTP to sponsor a second challenging round of tests, inviting participants to present predictions relative to the rodent carcinogenicity of a further 30 chemicals; these are currently being tested. In this article, we present our **predictions based on structure-activity relationship considerations**. In our procedure, first each chemical was assigned to an activity mechanism class and then, with semiquantitative considerations, was **assigned a probability carcinogenicity score, taking into account simultaneously the hypothesized action mechanism and physical chemical parameters.**
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9512429&form=6&db=m&Dopt=b

BENIGNI 97
*Mutat Res* 1997 Aug;387(1):35-45

The first US National Toxicology Program exercise on the prediction of rodent carcinogenicity: definitive results.

Benigni R

Istituto Superiore di Sanita, Laboratory of Comparative Toxicology and Ecotoxicology, Rome, Italy.

A few years ago, the US National Toxicology Program sponsored an exercise aimed at comparing different prediction approaches for carcinogenicity by challenging them on a common set of chemicals. The exercise was considered to be sufficiently completed when 40 (out of 44) chemicals were actually experimentally tested, and the experimental and estimated carcinogenicity were compared. More recently, the rodent results for the remaining 4 chemicals have been disclosed, making it possible to draw definitive conclusions on the comparative exercise. Having analyzed the first subset of results with multivariate statistical methods, we present here the analysis of the complete set of results. The present analysis also considers aspects (e.g., the **complementarity of the different systems in identifying the carcinogens**), which had not been investigated previously. The conclusion of this study were: (a) the expansion of the data base from 40 to 44 chemicals did not significantly change the results of the exercise; (b) the **structure-activity approaches generated prediction profiles different from those generated by the prediction systems mainly relying on the use of experimental data** (in vitro and in vivo); (c) the **performance of the predictive systems was generally rather limited**; (d) the **prediction systems were affected by over sensitivity**; they were generally capable of identifying the molecules containing the potentially alerting substructures, but were not so refined as to be able to discriminate between potential and actual carcinogenicity; (e) **the combination of the systems into batteries did not permit a significant increase in the performance of the individual methods**. The need for, and possible approaches to finely tuning the systems are discussed.

PMID: 9254891, UI: 97398677
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9254891&form=6&db=m&Dopt=b

BRISTOL 96
Bristol D.W., J.T. Wachsman, and A. Greenwell, 1996. The NIEHS Predictive-Toxicology Evaluation Project. Environmental Health Perspectives;**104** Suppl 5:1001-10.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933048&form=6&db=m&Dopt=b

Laboratory of Environmental Carcinogenesis and Mutagenesis, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA. bristol@niehs.nih.gov

The Predictive-Toxicology Evaluation (PTE) project conducts collaborative experiments that subject the performance of predictive-toxicology (PT) methods to rigorous, objective evaluation in a uniquely informative manner. Sponsored by the National Institute of Environmental Health Sciences, it takes advantage of the ongoing testing conducted by the U.S. National Toxicology Program (NTP) to estimate the true error of models that have been applied to make prospective predictions on previously untested, noncongeneric-chemical substances. The PTE project first identifies a group of standardized NTP chemical bioassays either scheduled to be conducted or are ongoing, but not yet complete. The project then announces and advertises the evaluation experiment, disseminates information about the chemical bioassays, and encourages researchers from a wide variety of disciplines to publish their predictions in peer-reviewed journals, using whatever approaches and methods they feel are best. A collection of such papers is published in this Environmental Health Perspectives Supplement, providing readers the opportunity to compare and contrast PT approaches and models, within the context of their prospective application to an actual-use situation. This introduction to this collection of papers on predictive toxicology summarizes the predictions made and the final results obtained for the 44 chemical carcinogenesis bioassays of the first PTE experiment (PTE-1) and presents information that identifies the 30 chemical carcinogenesis bioassays of PTE-2, along with a table of prediction sets that have been published to date. It also provides background about the origin and goals of the PTE project, outlines the special challenge associated with estimating the true error of models that aspire to predict open-system behavior, and summarizes what has been learned to date.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933048&form=6&db=m&Dopt=b

CUNNINGHAM 98
Cunningham AR, G. Klopman, H.S. Rosenkranz, 1998. Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutation Research* **31**; 405(1):9-27

A set of chemicals tested for carcinogenicity in rats that have been analyzed in the Carcinogenic Potency Database (CPDB) was subjected to CASE/MULTICASE (a computer-automated structure evaluation system) structure-activity relationship (SAR) analyses. This SAR system identifies structural features of chemicals in a learning set that are associated with a predefined activity and produces an SAR model based on these characteristics. The rat CPDB used in this study consisted of 745 chemicals, 383 of which are carcinogens, 14 marginally active carcinogens (i.e., chemicals that require a relatively high dose to induce carcinogenesis) and 348 are non-carcinogens. In an internal prediction analysis where CASE/MULTICASE 'predicted' the activity of chemicals in the learning set, the system was able to achieve a concordance between experimental and predicted results of 95%. This indicates that the program is able to adequately assess the chemicals in the database. **In a 10-fold cross-validation study where 10 disjoint sets of 10% of the chemicals were removed from the database and the remaining 90% of the chemicals were used as a learning set, CASE/MULTICASE was able to achieve a concordance between experimental and predicted results of 64%.** Using a modified validation process designed to investigate the predictivity of a more focused SAR model, the system was able to achieve a concordance of 71% between experimental and predicted results. Among the major biophores identified by CASE/MULTICASE as associated with cancer causation in rats, several are derived from electrophilic or potentially electrophilic compounds (e.g., aromatic amines, nitrogen mustards, isocyanates, epoxides). **Other biophores however are derived from chemicals seemingly devoid of actual or potential DNA-reactivity and as such may represent structural features of non-genotoxic carcinogens.**
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9729240&form=6&db=m&Dopt=b

ENSLEIN 90
*Mutagenesis* 1990 Jul;5(4):305-6
Prediction of probability of carcinogenicity for a set of ongoing NTP bioassays.
Enslein K, Blake BW, Borgstedt HH
Health Designs, Inc., Rochester, NY 14604.

Forty-four compounds currently undergoing carcinogenesis bioassay by the National Toxicology Program were submitted to the TOPKAT program for prediction of their potential carcinogenicity. Sixteen compounds could not be handled by TOPKAT. Of the 28 for which predictions were made, 26 (93%) had a confidence level in the estimate of at least moderate. Seventeen were predicted to be carcinogens and 11 non-carcinogens. These results will be compared with the assay results as the assays are completed.
PMID: 2398815, UI: 90376976
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=2398815&form=6&db=m&Dopt=b

JONES 96
Jones T.D. and C.E. Easterly, 1996. A RASH Analysis of National Toxicity Program Data: Predictions for 30 Compounds to Be Tested in Rodent Carcinogenesis Experiments. *Environmental Health Perspectives*; **104**S(5):1017-30
Health Sciences Research Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee

Relative potencies for 30 compounds scheduled for carcinogenic testing in the 2-year rodent bioassays were estimated based on comparisons with a wide variety of bioassay data for benzo[a]pyrene, nicotine, cisplatin, aflatoxin B1, and cyclophosphamide. Potential for oncogenic transformation of each of the compounds was estimated from short-term bioassays**. Promoting strength was assigned on the basis of comparisons of the product of relative potency and test dose with the distribution of similar products obtained for 67 common compounds** in the database of Gold et al. A potency class for promotion was assigned on the basis of whether the potency-adjusted test dosage was >2sigma below the mean, >1sigma below the mean, within +/-sigma of the mean, >sigma above the mean, or >2sigma above the mean, as determined from the 67 compounds. The underlying hypothesis is that a weak test dose may have a low probability of revealing a potential carcinogen, whereas a strong dose may have a high probability of producing false-positive results. Predictions are therefore directed at the central 68% of the log-normal frequency distribution according to the assumption that +/-sigma represents the ideal test dose.
PMID: 9512436
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9512436&form=6&db=m&Dopt=b

KING 96
Environ Health Perspect 1996 Oct;104 Suppl 5:1031-40
Prediction of rodent carcinogenicity bioassays from molecular structure using inductive logic programming.
King RD, Srinivasan A
Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, London, United Kingdom. rd_king@icrf.ac.uk

The machine learning program Progol was applied to the problem of forming the structure-activity relationship (SAR) for a set of compounds tested for carcinogenicity in rodent bioassays by the U.S. National Toxicology Program (NTP). Progol is the first inductive logic programming (ILP) algorithm to use a fully relational method for describing chemical structure in SARs, based on using atoms and their bond connectivities. **Progol is well suited to forming SARs for carcinogenicity as it is designed to produce easily understandable rules (structural alerts) for sets of noncongeneric compounds.** The Progol SAR method was tested by prediction of a set of compounds that have been widely predicted by other SAR methods (the compounds used in the NTP's first round of carcinogenesis predictions). For these compounds no method (human or machine) was significantly more accurate than Progol. **Progol was the most accurate method that did not use data from biological tests on rodents** (however, the difference in accuracy is not significant). The **Progol predictions were based solely on chemical structure and the results of tests for Salmonella mutagenicity.** Using the full NTP database, the prediction accuracy of Progol was estimated to be 63% (+/- 3%) using 5-fold cross validation. **A set of structural alerts for carcinogenesis was automatically generated and the chemical rationale for them investigated- these structural alerts are statistically independent of the Salmonella mutagenicity.** Carcinogenicity is predicted for the compounds used in the NTP's second round of carcinogenesis predictions. The results for prediction of carcinogenesis, taken together with the previous successful applications of predicting mutagenicity in nitroaromatic compounds, and inhibition of angiogenesis by suramin analogues, show that Progol has a role to play in understanding the SARs of cancer-related compounds.

PMID: 8933051, UI: 97087022

http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933051&form=6&db=m&Dopt=b

LEE 96
Environ Health Perspect 1996 Oct;104S(5):1059-63
Carcinogenicity Predictions for a Group of 30 Chemicals Undergoing Rodent Cancer Bioassays Based on Rules Derived from Subchronic Organ Toxicities.
Lee Y, Buchanan BG, Rosenkranz HS

Intelligent Systems Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania

**Rodent carcinogenicities** for a group of 30 chemicals which form the subject of the Second NIEHS Predictive-Toxicology Evaluation Experiment **are predicted based on their subchronic organ toxicities**. Predictions are made by rules learned by the rule learning (RL) induction program.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9512439&form=6&db=m&Dopt=b

LEWIS 98
Lewis D.F., C. Ioannides, D.V. Parke, 1998. Further evaluation of COMPACT, the molecular orbital approach for the prospective safety evaluation of chemicals. Mutation Research,**13**;412(1):41-54
School of Biological Sciences, University of Surrey, Guildford, UK.

The **molecular dimensions and electronic structures** of the first group of 100 US NCI/NTP miscellaneous chemicals, evaluated for potential carcinogenicity by computer-optimized molecular parametric analysis for chemical toxicity (**COMPACT**) have been re-determined. Using improved criteria for cytochrome P450 (CYP) substrate specificity, re-defined for CYP1 as having a COMPACT radius [square root of (deltaE - 9.5)2 + (a/d(2) - 7.8)2] of < 6.5, and for CYP2E as having a collision diameter of 6.5 angstroms or less and deltaE < 15.5, the likely substrates of CYP1 and CYP2E, which are regarded as potential carcinogens, have been identified. In addition, log P values have been taken into account; those chemicals with log P < 0 are non-lipophilic substrates unlikely to reach the activating cytochrome enzymes, and have been regarded as non-carcinogens. The second group of 100 US NCI/NTP chemicals have also now been categorized by COMPACT into CYP1 and CYP2E substrates, and their potential carcinogenicities evaluated. Of the 203 chemicals in the 2 groups, those positive in the rodent two-species life-span carcinogenicity study (rodent assay) were 53%, those positive in the Ames test (mutagenicity) were 48%, and those positive in the COMPACT programme (carcinogenicity, mutagenicity, cytotoxicity) were 54%. **Concordance between the COMPACT prediction of carcinogenicity/cytotoxicity and rodent two species life-span carcinogenicity data for the 203 chemicals is 69%, and correlation of COMPACT with Ames test data is 61%. The sensitivity of COMPACT for predicting rodent carcinogenicity is 72%, whereas the sensitivity of the Ames test for predicting carcinogenicity for the 203 chemicals was only 57%.** The degree (severity) of rodent carcinogenicity also showed correlation with the COMPACT predictive evaluations of the chemicals.

PMID: 9508363, UI: 98167788
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9508363&form=6&db=m&Dopt=b

MARCHANT 96
C.A. Marchant, 1996. Prediction of Rodent Carcinogenicity Using the DEREK System for 30 Chemicals Currently Being Tested by the National Toxicology Program. *Environmental Health Perspectives;* **104**S(5): 1065-73

LHASA UK, School of Chemistry, University of Leeds, Leeds, United Kingdom

DEREK is a **knowledge-based expert system** for the qualitative prediction of toxicity. The DEREK system has been used to predict the carcinogenicity in rodents of the 30 chemicals in the second National Toxicology Program (NTP) carcinogenicity prediction exercise. Seven of the chemicals were predicted to be carcinogens. For 23 chemicals, there was no evidence in the DEREK knowledge base to suggest carcinogenic activity. Supplementary data from a variety of sources have been evaluated by human experts to assess confidence in each DEREK prediction. These sources included standard toxicology reference texts, genotoxicity and subchronic toxicity assay results for each chemical, as well as Salmonella mutagenicity and carcinogenicity

data for close structural analogues. This process has led to the proposal of a number of improvements to the DEREK carcinogenicity knowledge base.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9512442&form=6&db=m&Dopt=b

MORIGUCHI 96
Environ Health Perspect 1996 Oct;104S(5):1051-8
Prediction of the Rodent Carcinogenicity of Organic Compounds from Their Chemical Structures Using the FALS Method.
Moriguchi I, Hirano H, Hirono S
School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan

**Fuzzy adaptive least-squares** (FALS), a pattern recognition method recently developed in our laboratory for correlating structure with activity rating, was used to generate quantitative structure-activity relationship (QSAR) models on the carcinogenicity of organic compounds of several chemical classes. Using the predictive models obtained from the chemical class-based FALS QSAR approach, the rodent carcinogenicity or noncarcinogenicity of a group of organic chemicals currently being tested by the U.S. National Toxicology Program was estimated from their chemical structures.

PMID: 9512443

PERROTTA 96
Perrotta A, D. Malacarne, M. Taningher, R. Pesenti, M. Paolucci, S. Parodi, 1996. A computerized connectivity approach for analyzing the structural basis of mutagenicity in Salmonella and its relationship with rodent carcinogenicity. *Environ Mol Mutagen*; **28**(1):31-50 http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8698045&form=6&db=m&Dopt=b
Perrotta A, Malacarne D, Taningher M, Pesenti R, Paolucci M, Parodi S
Laboratorio di Oncologia Sperimentale, Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy.

We have applied **a new software program, based on graph theory** and developed by our group, **to predict mutagenicity in Salmonella.** The software analyzes, as information in input, the structural formula and the biological activities of a relatively large database of chemicals to generate any possible molecular fragment with size ranging from two to ten nonhydrogen atoms, and detects (as predictors of biological activity) those fragments statistically associated with the biological property investigated. Our previous work used the program to predict carcinogenicity in small rodents. In the current work we applied a modified version of the program, which bases its predictions solely on the most important fragment present in a given molecule, considering as practically negligible the effects of additional less important fragments. For Salmonella mutagenicity we used a database of 551 compounds, and the program achieved **a level of predictivity (73.9%) comparable to that obtained by other authors using the Computer Automated Structure Evaluation (CASE) program**. We evaluated the relative contributions of biophores and biophobes to overall predictivity: biophores tended to be more important than biophobes, and chemicals containing both biophores and biophobes were more difficult to predict. **Many of the molecular fragments identified by the program as being strongly associated with mutagenic activity were similar to the structural alerts identified by the human experts Ashby and Tennant.** Our results tend to confirm that structural alerts useful to predict Salmonella mutagenicity are generally not very strong predictors of rodent carcinogenicity. Although the predictivity level achieved for oncogenic activity improved when the program was directly trained with carcinogenicity data, carcinogenicity as a biological endpoint was still more difficult to predict than Salmonella mutagenicity.

PMID: 8698045, UI: 96313172
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8698045&form=6&db=m&Dopt=b

PURDY 96
Environ Health Perspect 1996 Oct;104 Suppl 5:1085-94
A mechanism-mediated model for carcinogenicity: model content and prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 25 organic chemicals.
Purdy R

3M Environmental Laboratory, St Paul, Minnesota 55133, USA.

A hierarchical model consisting of quantitative structure-activity relationships based mainly on chemical reactivity was developed to predict the carcinogenicity of organic chemicals to rodents. The model is comprised of quantitative structure-activity relationships, **QSARs** based on hypothesized mechanisms of action, metabolism, and partitioning. **Predictors included octanol/water partition coefficient, molecular size, atomic partial charge, bond angle strain, atomic acceptor delocalizibility, atomic radical superdelocalizibility, the lowest unoccupied molecular orbital (LUMO) energy of hypothesized intermediate nitrenium ion of primary aromatic amines, difference in charge of ionized and unionized carbon-chlorine bonds, substituent size and pattern on polynuclear aromatic hydrocarbons, the distance between lone electron pairs over a rigid structure, and the presence of functionalities such as nitroso and hydrazine**. The **model correctly classified 96% of the carcinogens in the training set of 306 chemicals, and 90% of the carcinogens in the test set of 301 chemicals.** The test set by chance contained 84% of the positive thio-containing chemicals. A QSAR for these chemicals was developed. This posttest set modified model correctly predicted 94% of the carcinogens in the test set. This model was used to predict the carcinogenicity of the 25 organic chemicals the U.S. National Toxicology Program was testing at the writing of this article.

PMID: 8933058, UI: 97087029
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8933058&form=6&db=m&Dopt=b

TENNANT 96
Environ Health Perspect 1996 Oct;104S(5):1095-100
Predictions for the Outcome of Rodent Carcinogenicity Bioassays: Identification of Trans-species Carcinogens and Noncarcinogens.
Tennant RW, Spalding J
National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina

Thirty chemicals or substances currently undergoing long-term carcinogenicity bioassays in rodents have been used in a project to further evaluate methods and information that may have the capability of predicting potential carcinogens. In our predictions the **principal information used includes structural alerts and in vitro test results for Salmonella mutagenicity, relative subchronic toxicity, and the sites and types of pathology found in subchronic (90-day) studies**. This group of chemicals differs significantly from those used previously to evaluate predictive methods in that 23 of 30 are defined as nonmutagenic by conventional criteria. The goal of this predictive effort is to identify categorically the chemicals that have the capacity to induce cancers in both rats and mice (trans-species carcinogens) and those that are not carcinogenic in either rats or mice. Chemicals that show properties that may be associated with tumor induction in either species, i.e., **species-specific cancers, are categorized as being of "uncertain predictability."** This category includes chemicals believed to have limited carcinogenic potential that is manifested principally as a consequence of the genetic background of the test strain of inbred rodent.

PMID: 9512449
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9512449&form=6&db=m&Dopt=b

TENNANT 90
Tennant R.W., J. Spalding, S. Stasiewicz, J. Ashby, 1990. Prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 44 chemicals by the National Toxicology Program. *Mutagenesis*; **5**(1):3-14
Cellular and Genetic Toxicology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.

This paper was written to enable evaluation of the concept that **knowledge about chemical structure combined with limited short-term genotoxicity and toxicity test results** can **be used to predict potential carcinogens**. Previous attempts have been potentially biased by prior knowledge about the tumorigenicity of chemicals in animals or humans, but the 44 chemicals that are currently being bioassayed for carcinogenicity by the National Toxicology Program provide an opportunity prospectively to evaluate factors that may be predictive of chemical carcinogenicity. Predictions of rodent carcinogenicity for these 44 agents are presented as an example of what we believe is the best available approach at this time. This publication will also enable others to make their own predictions (using whatever methods they believe to have high predictive value) before the results of the animal assays are known.

RICHARD 98
Mutat Res 1998 May 25;400(1-2):493-507
Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet?
Richard AM
MD-68, Environmental Carcinogenesis Division, National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA. richard.ann@epamail.epa.gov

There is a great deal of current interest in the use of commercial, automated programs for the prediction of mutagenicity and carcinogenicity based on chemical structure. However, the goal of accurate and reliable toxicity prediction for any chemical, based solely on structural information remains elusive. The toxicity prediction challenge is global in its objective, but limited in its solution, to within local domains of chemicals acting according to similar mechanisms of action in the biological system; to predict, we must be able to generalize based on chemical structure, but the biology fundamentally limits our ability to do so. Available commercial systems for mutagenicity and/or carcinogenicity prediction differ in their specifics, yet most fall in two major categories: (1) automated approaches that rely on the use of statistics for extracting correlations between structure and activity; and (2) knowledge-based expert systems that rely on a set of programmed rules distilled from available knowledge and human expert judgement. These two categories of approaches differ in the ways that they represent, process, and generalize chemical-biological activity information. An application of four commercial systems (TOPKAT, CASE/MULTI-CASE, DEREK, and OncoLogic) to mutagenicity and carcinogenicity prediction for a particular class of chemicals-the haloacetic acids (HAs)-is presented to highlight these differences. Some discussion is devoted to the issue of gauging the relative performance of commercial prediction systems, as well as to the role of prospective prediction exercises in this effort. And finally, an alternative approach that stops short of delivering a prediction to a user, involving structure-searching and data base exploration, is briefly considered.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9685707&form=6&db=m&Dopt=b

SANDERSON 91
Hum Exp Toxicol 1991 Jul;10(4):261-73
Computer prediction of possible toxic action from chemical structure; the DEREK system.
Sanderson DM, Earnshaw CG
Schering Agrochemicals Limited, Chesterford Park Research Station, Saffron Walden, Essex, UK.

1. The development of **DEREK, a computer-based expert system** (derived from the LHASA chemical synthesis design program) for the qualitative prediction of possible toxic action of compounds on the basis of their chemical structure is described. 2. The system is able to perceive chemical sub-structures within molecules and relate these to a rulebase linking the sub-structures with likely types of toxicity. 3. Structures can be drawn in directly at a computer graphics terminal or retrieved automatically from a suitable in-house database. 4. The system is intended to aid the selection of compounds based on toxicological considerations, or separately to indicate specific toxicological properties to be tested for early in the evaluation of a compound, so saving time, money and some laboratory animals and resources.

PMID: 1679649, UI: 91363010

WILLIAMS 96
Williams G.M., M.J. Iatropoulos, J.H. Weisburger JH, 1996. Chemical carcinogen mechanisms of action and implications for testing methodology. *Experimental Toxicology and Pathology;* **48**(2-3):101-11.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8672863&form=6&db=m&Dopt=b
American Health Foundation, Valhalla, NY 10595, USA.

Chemical carcinogens are of two distinct types, DNA-reactive and epigenetic. Testing methodology can be directed toward detecting effects of both types of carcinogen. Carcinogens of the DNA-reactive type are defined by the formation of covalently bound DNA adducts. These chemicals have structures that yield electrophilic reactants either directly or after bioactivation. These agents cause genomic alteration in the structure or function of DNA in the target cell. In addition, these compounds can exert other cellular and tissue epigenetic effects, such as cell proliferation and growth promotion. Carcinogens of the epigenetic (paragenetic) type, in contrast, do not react with DNA, but rather display cellular effects such as neoplasm growth promotion, cytotoxicity, inhibition of tissue growth regulation, peroxisome proliferation, endocrine modification, immunosuppression and/or sustained tissue ischemia that can be the basis for increases in neoplasia. Their chemical structure is such that they do not give rise to a reactive electrophile. The testing methodologies to identify either type follow a Decision Point Approach designed to identify potential carcinogenicity and yield mechanistic information on the production of effects that underlie carcinogenicity. It has 5 stages focusing on the chemical structure, DNA-reactivity, epigenetic effects, limited bioassays and finally the application of the accelerated bioassay (ABA). ABA requires 40 weeks and applies the use of sensitive markers for induction of neoplasia in comparison to positive control compounds for important organs in human carcinogenesis. It enables data acquisition of the entire carcinogenic process directed toward developing mechanistic information. The ABA has the potential to replace the chronic bioassay in rodents in some circumstances and can serve as an alternative to a chronic bioassay in a second species.
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8672863&form=6&db=m&Dopt=b

WEISBURGER 96
Environ Health Perspect 1996 Oct;104S(5):1105-12
Multicomponent Criteria for Predicting Carcinogenicity: Dataset of 30 NTP Chemicals.
Huff J, Weisburger E, Fung VA
National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina

This article is in response to the challenge issued to the scientific community by the National Toxicology Program to predict the carcinogenicity potential of 30 chemicals previously selected for long-term carcinogenicity testing. Utilizing the available toxicologic, genetic, and structural information on 30 chemicals previously selected for long-term carcinogenicity testing, we predict that 16 chemicals (53%) would induce some indication of carcinogenic activity in rodents; we further predict that 10 chemicals (33%) would be associated with weak or equivocal carcinogenic responses, and another 4 (13%) would give no indication of carcinogenicity. Our level of certainty is indicated for many of these predictions. Nonetheless, we believe that most instances of guessing whether a chemical would eventually induce cancer in experimental animals and hence represent a carcinogenic hazard to humans are fraught with considerable uncertainty: uncertainty that can only be relieved by long-term testing for carcinogenicity in animals or by conducting an epidemiologic investigation of exposed individuals or groups. We further believe that the day may come when our predictive acumen will be upgraded to such an extent that we might eventually obviate cancer testing. Until then, and in the best interests of public health, however, we urge long term testing of chemicals in animals be continued, at increased pace.

WOO 95
Toxicol Lett 1995 Sep;79(1-3):219-28
Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals.
Woo YT, Lai DY, Argus MF, Arcos JC
Health and Environmental Review Division (7403), U.S. Environmental Protection Agency, Washington, D.C. 20460, USA.

Since the inception of Section 5 (Premanufacturing/Premarketing Notification, PMN) of the Toxic Substances Control Act (TSCA), structure-activity relationship (SAR) analysis has been effectively used by U.S. Environmental Protection Agency's (EPA) Structure Activity Team (SAT) in the assessment of potential carcinogenic hazard of new chemicals for which test data are not available. To capture, systematize and codify the Agency's predictive expertise in order to make it more widely available to assessors outside the TSCA program, a cooperative project was initiated to develop a knowledge rule-based expert system to mimic the thinking and reasoning of the SAT. In this communication, we describe the overall structure of this expert system, discuss the scientific bases and principles of SAR analysis of chemical carcinogens used in the development of SAR knowledge rules, and delineate the major factors/rules useful for assessing the carcinogenic potential of fibers, polymers, metals/metalloids and several major classes of organic chemicals. An integrative approach using available short-term predictive tests and non-cancer toxicological data to supplement SAR analysis has also been described.

YE 98
*Journal of the American Statistical Association,* Volume 93, Number 441, March 1998, pp. 120–131
On Measuring and Correcting the Effects of Data Mining and Model Selection
Jianming Ye

In the theory of linear models, the concept of degrees of freedom plays an important role. This concept is often used for measurement of model complexity, for obtaining an unbiased estimate of the error variance, and for comparison of different models. I have developed a concept of generalized degrees of freedom (GDF) that is applicable to complex modeling procedures. The definition is based on the sum of the sensitivity of each fitted value to perturbation in the corresponding observed value. The concept is nonasymptotic in nature and does not require analytic knowledge of the modeling procedures. The concept of GDF offers a unified framework under which complex and highly irregular modeling procedures can be analyzed in the same way as classical linear models. By using this framework, many difficult problems can be solved easily. For example, one can now measure the number of observations used in a variable selection process. Different modeling procedures, such as a tree-based regression and a projection pursuit regression, can be compared on the basis of their residual sums of squares and the GDF that they cost. I apply the proposed framework to measure the effect of variable selection in linear models, leading to corrections of selection bias in various goodness-of-fit statistics. The theory also has interesting implications for the effect of general model searching by a human modeler.
http://www.amstat.org/publications/jasa/abstracts_98/YE.HTM

YOSHIKAWA 96
Yoshikawa, K., 1996. Anomalous nonidentity between Salmonella genotoxicants and rodent carcinogens: nongenotoxic carcinogens and genotoxic noncarcinogens. Environmental Health Perspectives; **104**(1):40-6
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8834860&form=6&db=m&Dopt=b\
Yokohama Research Center, Mitsubishi Chemical Corporation, Japan.

According to current data, the **capacity to cause nonprogrammed or unscheduled cell proliferation in target tissues, a common characteristic of chemical carcinogens, may play a more important role in the development of tumors than does genotoxicity**. This paper provides strong support for the validity of this conclusion. Ames-negative nongenotoxicants may be considered to be carcinogenic primarily because of their ability to induce cell proliferation in animal tissues and organs. In addition, such nongenotoxic carcinogens may also provide latent and modest DNA (equivocal) modifications that never lead to Ames-positive events. Conversely, noncarcinogenesis by Ames-positive agents is likely to be linked to a lack of stimulation of cell division. Nongenotoxic and genotoxic carcinogens rely on both cell proliferation and equivocal DNA modification for their full carcinogenicity. Such equivocal DNA modifications do not appear to be formed by tumor promoters. The role of cell proliferation may provide a favorable milieu for the occurrence of genetic instability, give rise to selective "apoptosis-resistant abnormal cells," and then affect clonal expansion of these cells. Therefore, understanding the influence of nongenotoxic and genotoxic carcinogens on cell proliferation capability is a key point in determining the mechanisms of chemical carcinogenesis. Considering the contradictory and common features of genotoxicants and carcinogens, **early detection of nonprogrammed cell proliferation is the most effective approach to predict human and rodent carcinogenicity.**
http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=8834860&form=6&db=m&Dopt=b

KNOWLEDGESEEKER™

KnowledgeSEEKER Product Information (http://www.angoss.com/ksprod/kspage.htm)

KnowledgeSEEKER is a data mining software tool that uses a unique cross-referencing process that enables businesses to draw conclusions from varied and disparate databases. The application of KnowledgeSEEKER can be tailored to suit the specific needs of any number of different business tasks, from customer profiling and segmentation to fraud detection and risk analysis. The conclusion that may be drawn from the data can often be surprising, since the software works independent of any user bias, putting less importance on preconceived ideas.

KnowledgeSEEKER benefits from a simple, intuitive GUI (graphical user intreface) that is very easy to learn. It is faster and easier to use and interpret than both statistical models and new technologies such as neural networks. Analysis results are rapidly displayed in the form of a clear and interactive decision tree. In just a few minutes it examines all the relationships between the fields in your data-eliminating trial-and-error guesswork by searching and finding strong statistical relationships. All fields and combinations of problem are ranked in order of importance. Both the sensitivity of the correlation finding the volume of the information displayed are easily user-defined.

David W. Aha
http://www.aic.nrl.navy.mil/~aha/research/machine-learning.html
Some references on decision trees (missing many)
Compiled quickly on 12/7/94

1. Decision trees (just a small sampling of papers; missing is work over the years by Norton, by Catlett, by Utgoff, by Bratko, etc):


Breiman,~L., Friedman,~J.~H., Olshen,~R.~A., \& Stone,~C.~J. (1984). {\it Classification and regression trees.} Belmont, CA: Wadsworth International Group.

Buntine,~W., \& Niblett,~T. (1992). A further comparison of splitting rules for decision-tree induction. {\it Machine Learning}, {\it 8}, 75--86.

Cardie,~C. (1993). Using decision trees to improve case-based learning. In {\it Proceedings of the Tenth International Conference on Machine Learning} (pp. 25--32). Amherst, MA: Morgan Kaufmann.

Cestnik,~B., \& Bratko,~I. (1991). On estimating proabilities in tree pruning. In {\it Proceedings of the Fifth European Working Session on Learning} (pp. 138--150). Porto, Portugal: Springer-Verlag.

Cestnik,~B., Kononenko,~I., \& Bratko,~I. (1987). ASSISTANT-86: A knowledge-elicitation tool for sophisticated users. In I.~Bratko \& N.~Lavra\v{c} (Eds.), {\it Progress in machine learning}. Bled, Yugoslavia: Sigma Press.

Fayyad,~U.~M., \& Irani,~K.~B. (1990). What should be minimized in a decision tree? In {\it Proceedings of the Eighth National Conference on Artificial Intelligence} (pp. 749--754). Boston, MA: AAAI Press.

Fayyad,~U.~M., \& Irani,~K.~B. (1992). On the handling of continuous-valued attributes in decision tree generation. {\it Machine Learning}, {\it 8}, 87--102.

Fayyad,~U.~M., \& Irani,~K.~B. (1992). The attribute selection problem in decision tree generation. In {\it Proceedings of the Tenth National Conference on Artificial Intelligence} (pp. 104--110). San Jose, CA: AAAI Press.

Fayyad,~U.~M., \& Irani,~K.~B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In {\it Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence} (pp, ?--?). Chambery, France: Morgan Kaufmann.

Fayyad,~U.~M., Weir,~N., \& Djorgovski,~S. (1993). SKICAT: A machine learning system for automated cataloging of large scale sky surveys. In {\it Proceedings of the Tenth International Conference on Machine Learning} (pp. 112--119). Amherst, MA: Morgan Kaufmann.

Iba,~W., \& Langley,~P. (1992). Induction of one-level decision trees. In {\it Proceedings of the Ninth International Conference on Machine Learning} (pp. 233--240). Aberdeen, Scotland: Morgan Kaufmann.

Langley,~P., \& Sage,~S. (1994). Oblivious decision trees and abstract cases. In D.~W.~Aha (Ed.), {\it Case-Based Reasoning: Papers from the 1994 Workshop} (Technical Report WS-94-01). Menlo Park, CA: AAAI Press.

Mingers,~J. (1989). An empirical comparison of pruning methods for decision tree induction. {\it Machine Learning}, {\it 4}, 227--243.

Mingers,~J. (1989). An empirical comparison of selection measures for decision-tree induction. {\it Machine Learning}, {\it 3}, 319--342.

Murthy,~S., Kasif,~S., Salzberg,~S., \& Beigel,~R. (1993). OC1: Randomized induction of oblique decision trees. In {\it Proceedings of the Eleventh National Conference on Artificial Intelligence} (pp. 322--327). Washington, DC: AAAI Press. - see recent JAIR article also.

Niblett,~T., \& Bratko,~I. (1986). Learning decision rules in noisy domains. In {\it Proceedings of Expert Systems 1986} (pp. 25--34). Cambridge, England: Cambridge University Press.

Quinlan,~J.~R. (1979). Discovering rules by induction from large numbers of examples: a case study. In D.~Michie (Ed.), {\it Expert Systems in the Micro-Electronic Age. Edinburgh, Scotland: Edinburgh University Press.

Quinlan,~J.~R. (1983). Learning efficient classification procedures and their application to chess end games. In R.~S.~Michalski, J.~G.~Carbonell, \& T.~M.~Mitchell (Eds.), {\it Machine learning: An artificial intelligence approach}. San Mateo, CA: Morgan Kaufmann.

Quinlan,~J.~R. (1986). Induction of decision trees. {\it Machine Learning}, {\it 1}, 81--106.

Quinlan,~J.~R. (1986). The effect of noise on concept learning. In R.~S.~Michalski, J.~G.~Carbonell, \& T.~M.~Mitchell (Eds.), {\it Machine learning: An artificial intelligence approach} (Vol. II). San Mateo, CA: Morgan Kaufmann.

Quinlan,~J.~R. (1987). Decision trees as probabilistic classifiers. In {\it \mlwfour} (pp. 31--37). Irvine, CA: Morgan Kaufmann.

Quinlan,~J.~R. (1987). Generating production rules from decision trees. In {\it Proceedings of the Tenth International Joint Conference on Artificial Intelligence} (pp. 304--307). Milan, Italy: Morgan Kaufmann.

Quinlan,~J.~R. (1988). An empirical comparison of genetic and decision-tree classifiers. In {\it Proceedings of the Fifth International Conference on Machine Learning} (pp. 135--141). Ann Arbor, MI: Morgan Kaufmann.

Quinlan,~J.~R. (1989). Unknown attribute values in induction. In {\it Proceedings of the Sixth International Workshop on Machine Learning} (pp. 164--168). Ithaca, NY: Morgan Kaufmann.

Quinlan,~J.~R. (1993). Combining instance-based learning and model-based learning. To appear in {\it Proceedings of the Tenth International Conference on Machine Learning.} Amherst, MA: Morgan Kaufmann.

Quinlan,~J.~R. (1993). {\it C4.5: Programs for machine learning}. San Mateo, CA: Morgan Kaufmann.

Schaffer,~C. (1991). When does overfitting decrease prediction accuracy in induced decision trees and rule sets? In {\it Proceedings of the European Working Session on Learning} (pp. 192--205). Porto, Portugal: Springer-Verlag.

Schaffer,~C. (1992). Deconstructing the digit recognition problem. In {\it Proceedings of the Ninth International Conference on Machine Learning} (pp. 394--399). Aberdeen, Scotland: Morgan Kaufmann.

Schaffer,~C. (1993). Overfitting avoidance as bias. {\it Machine Learning}, {\it 10}, 113--152.

Utgoff,~P.~E. (1989). Incremental induction of decision trees. {\it Machine Learning}, {\it 4}, 161--186.

Utgoff,~P.~E., \& Brodley,~C.~E. (1990). An incremental method for finding multivariate splits for decision trees. In {\it Proceedings of the Seventh International Conference on Machine Learning} (pp. 58--65). Austin, TX: Morgan Kaufmann.

Watanabe,~L., \& Rendell,~L. (1991). Learning structural decision trees from examples. To appear in {\it Proceedings of the Twelfth International Joint Conference on Artificial Intelligence}. Sydney, Australia: Morgan Kaufmann.

2. Constructing Boolean combinations of test for decision trees:
(this is only from 1989-1991; there may be more since that time)

Matheus,~C.~J., \& Rendell,~L.~A. (1989). Constructive induction on decision trees. In {\it Proceedings of the Eleventh International Joint Conference on Artificial Intelligence} (pp. 645--650). Detroit, MI: Morgan Kaufmann.

Matheus,~C.~J. (1990). Adding domain knowledge to SBL through feature construction. In {\it Proceedings of the Eighth National

Conference on Artificial Intelligence} (pp. 803--808).
Boston, MA: AAAI Press.

Matheus,~C.~J. (1989). {\it Feature construction: An analytic framework and an application to decision trees.} Doctoral dissertation, Department of Computer Science, University of Illinois at Urbana-Champaign, Illinois.

Pagallo,~G. (1989). Learning DNF by decision trees. In {\it Proceedings of the Eleventh International Joint Conference on Artificial Intelligence} (pp. 639--644). Detroit, MI: Morgan Kaufmann.

Pagallo,~G., \& Haussler,~D. (1990). Boolean feature discovery in empirical learning. {\it Machine Learning}, {\it 5}, 71--100.

Pagallo,~G. (1990). {\it Adaptive decision tree algorithms for learning from examples.} Doctoral dissertation, Department of Computer Science, University of California at Santa Cruz, California.

Yang,~D., Blix,~G., \& Rendell,~L.~A. (1991). The replication problem: A constructive induction approach. In {\it Proceedings of the European Working Session on Learning} (pp. 44--61). Proto, Portugal: Springer-Verlag.

Yang,~D., Blix,~G., \& Rendell,~L.~A. (1991). A scheme for feature construction and a comparison of empirical methods. In {\it Proceedings of the Twelvth International Joint Conference on Artificial Intelligence} (pp. 699--704). Sydney, Australia: Morgan Kaufmann.

Yang,~D. (1991). {\it Feature discovery in decision tree representation. Unpublished master's thesis, University of Illinois at Urbana-Champaign, Illinois.

See also more recent work by Salzberg and his colleagues on OC2.
A recent paper has been published on their work in JAIR.

Another recent paper appeared recently in MLj by Brodley and Utgoff.