

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Reliability and Evaluation of Identification Models Exemplified by a Histological Diagnosis Model

L.W.D. van Raamsdonk, S. van der Vange,
M. Uiterwijk and M. J. Groot

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/51362>

1. Introduction

Expert systems, or more precise decision support systems, are valuable tools for structuring the results of scientific research and to translate this to knowledge. The decision support system Determinator is now used for several years as a platform for models to identify subjects [1, 2]. The system is based on the two main different procedures for identification [3, 4]; a single access key (tree) and a free access key (matrix). The latter option provides the possibility to calculate the match between the subject as chosen by the user and the objects as included in the data model, based on a range of characteristics. In addition, a matrix allows to make selections, to filter the set of available objects and to compare two objects for their variability.

Datamodels for Determinator can be constructed using a Developer, which is part of the entire Determinator platform. Besides defining the objects (descriptions, illustrations and labels), the characteristics, and the connection between them (the matrix), the Developer also allows to evaluate the structure of the data model. Several parameters and metadata for the evaluation of a data model are part of the Developer.

This chapter provides the logic basis for the Determinator platform and introduces the background and calculation of four different parameters for the evaluation of data models: the coverage of variability space of the total data model or of a single object, the redundancy in a data model, and the capability to distinguish between different objects. The way in which these parameters are developed and applied will be demonstrated using a real case concerning the diagnosis of illegal hormone treatment of veal calves [5, 6]. The applicability of the parameters will be discussed and the development of a specific case (histological diagnosis) in a general platform (Determinator) will be evaluated.

2. Material and methods

Conventions

A datamodel developed in the framework of the DSS Determinator includes the following tables:

- List of features, with image file names and descriptions,
- Groups of features, with names and descriptions,
- List of targets, with image file names, descriptions and labels,
- Match table, with the feature on the rows and the targets on the columns,
- Tree information per node, with descriptions and image file names.

A data model consists of n features (denoted by i, j) to describe m objects (targets in the terminology of Determinator, denoted by p, q, r, s). Every feature consists of two or more feature states (k, l , and K for composite features). The basic principles are defined using first order logic [4, 7].

2.1. Logic basis

Free access key

Every cell in the matrix *Target x Features* contains a decision rule. These decision rules describe the logical relationship between the feature states and the targets, by specifying valid feature states for each target. A feature state can apply to one or more targets which imply that there might be no unique relation:

$$F_{i,k} \Rightarrow \{T_p \vee \dots \vee T_s\} \quad (1)$$

with $F_{i,k}$ as feature state k of the i^{th} feature, and $T_p \dots T_s$ as a series of targets which can be assigned individually. Otherwise, applying more specific feature states can limit the choice of targets:

$$F_{i,k} \wedge F_{j,l} \Rightarrow T_p \quad (2)$$

with $F_{i,k}$ as feature state k of the i^{th} feature, $F_{j,l}$ as feature state l of the j^{th} feature, and T_p as target.

The use of different states of a feature can add to the separation capability of that feature. Assuming three feature states:

$$\{(F_{i,1} \vee F_{i,2} \Rightarrow p) \wedge (F_{i,2} \vee F_{i,3} \Rightarrow q)\} \Rightarrow (F_{i,2} \Rightarrow p \vee q) \quad (3)$$

In this logic distribution, feature state $F_{i,1}$ identifies exclusively target p , and feature state $F_{i,3}$ identifies exclusively target q , but feature state $F_{i,2}$ can either identify target p or target q . This dual relationship can be indicated as *overlap* in a Venn diagram.

The DSS Determinator allows the user to choose a subject for identification and to answer a range of questions denoting the n features available in the model. Every possible answer represents a certain feature state k . The match between the chosen subject (represented by the answers given) and a target p is calculated by summing up all the true relationships between the chosen feature states $F_{i,k}$ and the defined target p :

$$\sum_{i=1}^n ([F_{i,k} \Rightarrow T_p] * W_i) \tag{4}$$

with W_i as weighting factor for feature i . The sums for all targets are represented as *Match percentages* in the output of the system and listed in descending order.

Single access key

A typical dichotomous tree consists of nodes (lemmas), which can point to either two targets (leaves), two nodes (branches) or combination of the two. Basically, every lemma in a tree is based on the decision rule:

$$F_{i,K} \Rightarrow P(x), \quad \neg F_{i,K} \Rightarrow \neg P(x) \tag{5}$$

The functions $P(x)$ and $\neg P(x)$ can describe a target or a further node. The structure of a dichotomous key can be defined as:

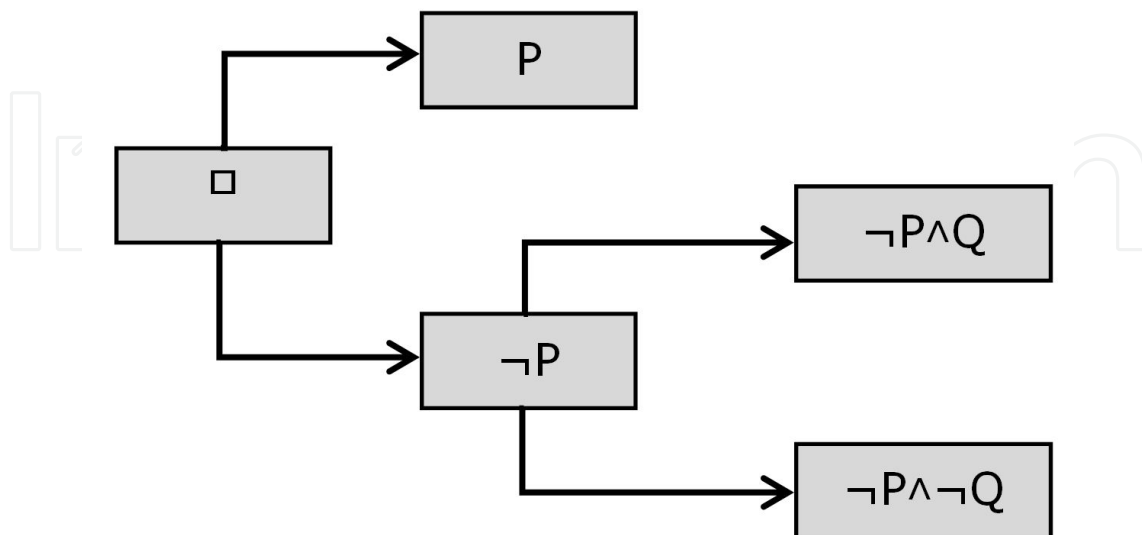


Figure 1. A hypothetical tree with two features indicated by the functions P and Q and three targets.

The combined feature state $F_{i,K}$ can combine more than one simple state, e.g. k and l . Determinator allows to construct a tree in which a node can point to a node in another part of the tree, and more than one node can point to a defined target T_p .

2.2. Quality parameters

The following parameters for validation of data models are being developed and evaluated in the framework of this paper.

Redundancy

Overlap between the areas of two targets exists when a variability range for target p overlap with the variability range for target q for the same feature (see figure 2, targets B and C; equation (3)). The overlap between the area of target p and of target q is the sum of the overlap regions for all features. Assuming the set of feature states that apply to target p as $\{F_{i,pmin}, F_{i,pmax}\}$ and the set of feature states that apply to target q as $\{F_{i,qmin}, F_{i,qmax}\}$, then:

$F_{i,a}$ = smallest $\{F_{i,pmax}, F_{i,qmax}\}$ (upper limit of the overlap region)

$F_{i,b}$ = largest $\{F_{i,pmin}, F_{i,qmin}\}$ (lower limit of the overlap region)

min_i = smallest $\{F_{i,pmin}, F_{i,qmin}\}$ (lower limit of the feature state range of both targets)

max_i = largest $\{F_{i,pmax}, F_{i,qmax}\}$ (upper limit of the feature state range of both targets)

Overlap per feature:

$$r_{i,p,q} = \frac{F_{i,p} - F_{i,q} + 1}{max_i - min_i + 1}; r_{i,p,q} \geq 0 \quad (6)$$

Average overlap for all feature differences between two targets p and q :

$$R_{p,q} = \left(\sum_{i=1}^n \frac{F_{i,p} - F_{i,q} + 1}{max_i - min_i + 1} \right) / n \quad (7)$$

The average redundancy of the total data model is the averaged overlap of every combination between two targets p and q . There are $(m * (m - 1)) / 2$ different combinations of targets.

Average redundancy:

$$R_{tot} = \left(\sum_{p=1}^m \left(\sum_{q=1}^{p-1} R_{p,q} \right) \right) / \frac{m * (m - 1)}{2} * 100 \quad (8)$$

The smaller the average redundancy, the smaller the chance that a certain range of feature states of a chosen subject will result in two or more match percentages of 100 % (according to (4); see object 3 in Figure 2). Redundancy is related to the correlation coefficients among features.

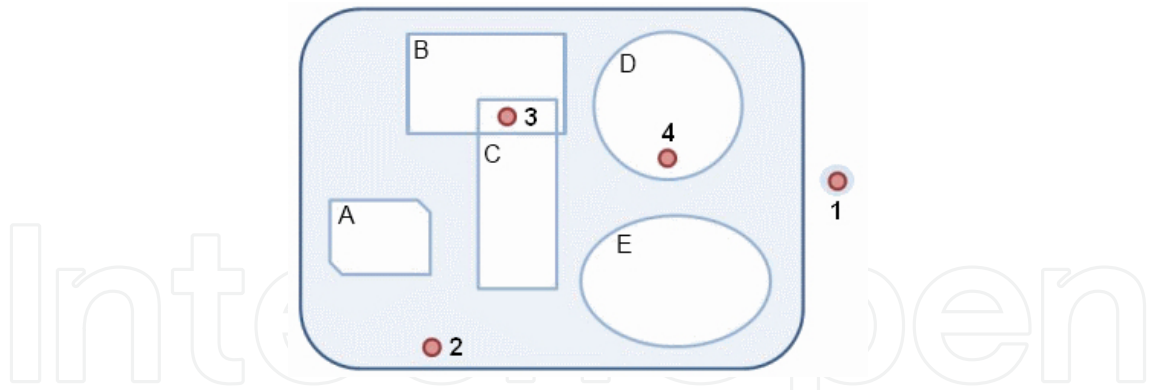


Figure 2. A hypothetical variation space with five targets A-E and four user chosen subjects. 1: subject outside the total variation space of the data model, a 100% match is impossible; 2: subject inside the total variation space of the data model, but without fit with one of the targets, a 100% match will not occur; 3: subject in the overlap of the variation of two or more targets, two or more 100% matches will result; 4: subject in one and only one variation space of a target, one 100% match will be found.

Uniqueness

The capability to distinguish between two targets p and q depends on the presence of at least one feature with unique variability ranges for each of the two targets. If overlapping regions exist for all features, there is at least a possibility to have a set of features states, describing a chosen subject, which shows a full match with more than one target. So, two targets p and q can uniquely be differentiated if and only if a feature i exists for which no state identifies target p as well target q :

$$U_{p,q} \Leftrightarrow \exists i : P(\forall k : Q(F_{i,k} \Rightarrow T_p \wedge F_{i,k} \not\Rightarrow T_q)) \quad (9)$$

This can be rewritten as:

$$U_{p,q} \Leftrightarrow \exists i : P(r_{i,p,q} = 0) \quad (9b)$$

with: $U_{p,q} = \text{TRUE}$: the two targets p and q have at least for one feature i non-overlapping feature ranges; there is at least one value $r_{i,p,q}$ equalling zero (equation (6)), and there is at least one feature indicated red in the menu option Compare of Determinator.

$U_{p,q} = \text{FALSE}$: the two targets p and q have overlapping ranges for all features; there is no value $r_{i,p,q}$ equalling zero (equation (6)), and there is no feature indicated red in the menu option Compare of Determinator.

If the distinction between two targets is based on only one feature i with a value $r_{i,p,q}$ equalling zero (no overlap), then the distinction could be considered as weak. Targets A, C and E in Figure 2 can be distinguished along the X-axis, targets A and B, and targets D and E can be distinguished along the Y-axis, whereas targets B and C can neither be distinguished

along the X-axis nor the Y-axis. If the group with the distinctive feature is disabled, Determinator could give for more than one target a full match in a query.

Separation capability

A data model can identify uniquely every target if and only if every combination between two targets p and q can be described with $U_{p,q} = \text{TRUE}$:

$$D_{tot} = \left(\sum_{p=1}^m \left(\sum_{q=1}^{p-1} U_{p,q} = \text{TRUE} \right) \right) / \frac{m*(m-1)}{2} * 100 \quad (10)$$

A datamodel can be indicated as suboptimal or not valid when the differentiation coefficient D_{tot} is less than 100%. Whether or not a data model could be validated with a value for D_{tot} lower than 100 % depends on the intention to have non-distinguishable targets (synonyms) present in the model or not.

Coverage of variability space

Every target possesses a part of the n -dimensional space defined by the data model. The share of a target in the total space is calculated as:

Coverage of space of a single target p :

$$o_p = \left(\prod_{i=1}^n (s_{i,p} / t_i) \right) * 100 \quad (11)$$

with:

$$s_{i,p} = F_{i,pmax} - F_{i,pmin} + 1$$

$$t_i = F_{i,max} - F_{i,min} + 1$$

In the situation of D_{tot} equalling 100 %, the sum of all individual target coverages is an indication of the total coverage of the variability space:

$$O_{tot} = \sum_{p=1}^m o_p ; D_{tot} = 100\% \quad (12)$$

The larger the coverage of the total variability space, the smaller the chance that a certain range of values of a subject will result in no match percentage of 100 % (according to (4)). In the situation that D_{tot} is smaller than 100% an overestimation occurs.

The diagnosis for illegal growth hormone use in veal calves will be used as illustration of model development and performance testing.

3. Model development and application

The use of illegal growth promoters is, although prohibited in the European Union, still part of current practice in animal farming. Reasonable monitoring of the hormones is hampered by the fact that the hormone or hormone cocktail is metabolised or excreted within a period of a few weeks. The effects of the use of hormones, however, can be seen in histological stained sections of either the prostate (male calves) or gland of Bartholin (female calves) with different staining techniques. The monitoring by means of histological examinations appears to be an important instrument in maintaining legislation for food safety and animal health [5, 6]. The interpretation of histological disorders needs a high level of expertise. An expert model has been developed in the framework of the DSS Determinator, in order to support the user to identify the extent of hormone treatment of veal calves. The different quality parameters will be illustrated after a further presentation of the model.

The data model consists of 13 features to identify a treatment level indicated as “normal”, “suspect” or “positive”. The features are presented in Table 1, and some of them are illustrated in Figure 3.

Group	Number	Feature	States
I	1	Presence of metaplasia (male)	[none,mild,severe]
II	2	Ratio between ducts and glandular tissue	[normal,more_ducts,mainly_ducts]
II	3	Presence of metaplasia (female)	[none,mild,severe]
II	4	Combined presence of metaplasia and an elevated duct ratio	[no,yes]
III	5	Presence of hyperplasia	[none,mild,severe]
III	6	Presence of cysts	[none,mild,severe]
III	7	Presence of hypersecretion	[none,mild,severe]
III	8	Presence of vacuolisation	[none,mild,severe]
III	9	Presence of mucus cells	[none,moderate,severe]
III	10	Presence of inflammation	[none,moderate,severe]
III	11	Presence of folding in the urethra	[none,moderate,severe]
III	12	Presence of thickening in the urethra	[none,moderate,severe]
IV	13	Number of deviating features	[0,.....,9]

Table 1. Overview of histological features for identifying the level of hormone treatment. The number of deviating features (group IV) is the sum of features in group III and either group I (male) or group II (female) that have a state differing from “none”. Feature 2 is excluded from this sum since it only applies to female animals.

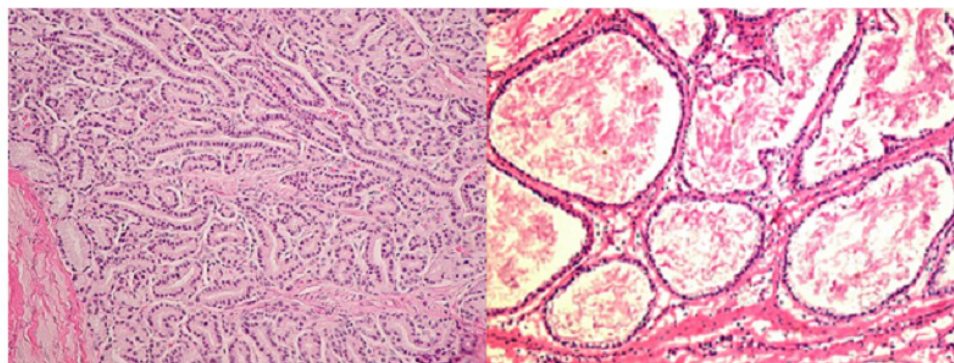


Figure 3. Normal appearance of prostate (left) and intensive presence of hypersecretion and cysts (right).

There are two strategies to reach a diagnosis:

- A. A quick, general diagnosis. Depending on the sex of the calf, selecting either feature groups I and IV (male) or groups II and IV (female) is sufficient.
- B. An extended diagnosis. In addition to the feature groups as indicated in strategy A group III is necessary.

The kernel of the data model consists of the groups I, II and IV to give a diagnosis of the treatment level. The diagnosis for possible hormone treatment in female calves is more complicated than for male calves. This is caused by the natural production of oestrogen hormones, which is lacking in male calves. The simple diagnosis <IF metaplasia=present THEN target positive> needs further support in female calves. A second diagnostic feature is used based on a larger share of ducts in the glandular tissue. The basic rule is then expanded to <IF metaplasia=present AND duct_ratio=elevated THEN target positive>. For both male and female calves the diagnosis “suspect” is supported by the number of deviating features. The duct ratio is excluded from this feature since it applies only to female calves. The logic tables to diagnose the level of treatment are presented in Table 2.

female	Metaplasia	
	[mild,severe]	[none]
	[more,mainly]	“positive”
[normal]	#=[1,2] → “normal”	#=[0,1,2] → “normal”
	#=[3,...,9] → “suspect”	#=[3,...,8] → “suspect”
male	Metaplasia	
	[mild,severe]	[none]
	“positive”	#=[0,1,2] → “normal”
		#=[3,...,8] → “suspect”

Table 2. Logic tables for the diagnosis of hormone treatment in veal calves. #: total number of deviating features including the presence of metaplasia, excluding an elevated duct ratio.

The diagnoses as illustrated in Table 2 can be extended further by including the individual features of group III (Table 1). The number of deviating features (feature 13) needs to be adjusted accordingly. The basic rules are translated in a formal decision tree, as shown in Figure 4.

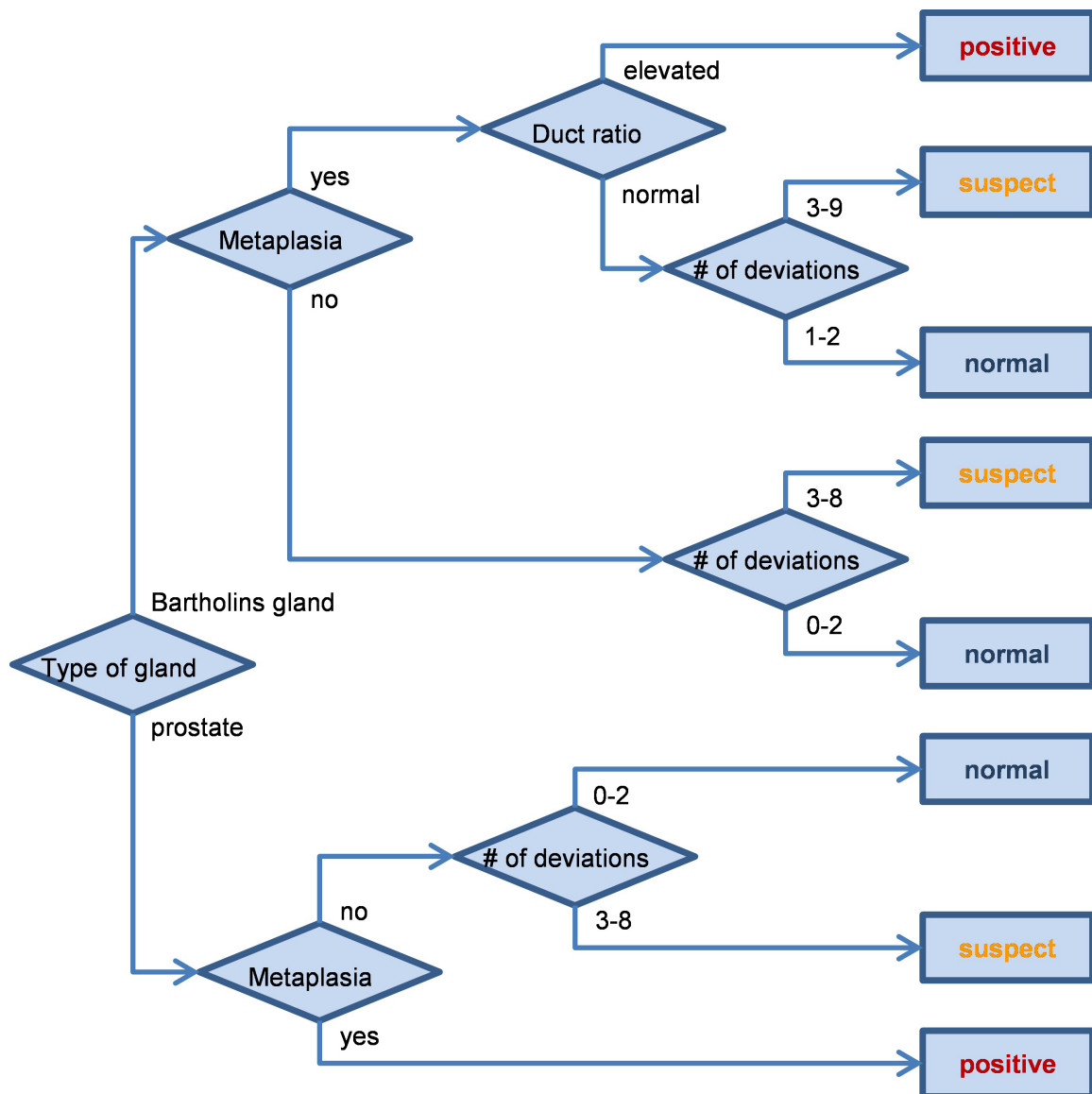


Figure 4. Decision tree for the diagnosis of hormone treatment in veal calves.

Finally, the decision tree is used as basis for a free access key. The importance and position of the feature indicating the presence of metaplasia is different for male and female diagnosis. For the latter only the combination of metaplasia and elevated duct ratio is decisive for the diagnosis “positive”. As a consequence, the presence of metaplasia is included twice in the free access key as feature 1 (group I for male animals) and feature 3 (group II for female animals). The free access key was optimised by giving all features a suitable weighting factor. All features of group III got the factor one.

The performance of the model is tested in eight runs following the two strategies. The continuous feature 13 is varied between 0 and 9 in every run in combination with the appropriate choices for the other features, as follows:

A1 (male): groups I and IV are used. Choice for feature 1 is [none].

A2 (male): groups I and IV are used. Choice for feature 1 is [mild] unless:

$$F_{13,k}=0 \rightarrow F_{1,k}=[\text{none}].$$

A3 (female): groups II and IV are used. Choices for features 2 and 3 are [normal] and [none].

A4 (female): groups II and IV are used. Choices for features 2 and 3 are [more_ducts] and [none].

A5 (female): groups II and IV are used. Choices for features 2 and 3 are [more_ducts] and [mild] unless:

$$F_{13,k}=0 \rightarrow F_{3,k}=[\text{none}].$$

B1 (male): groups I, III and IV are used. Choice for features 1 is [none].

B2 (male): groups I, III and IV are used. Choice for features 1 is [mild] unless:

$$F_{13,k}=0 \rightarrow F_{1,k}=[\text{none}].$$

$F_{13,k} > 1 \rightarrow$ the appropriate number of features of group B and C get the state [mild] or [moderate].

B3 (female): groups II, III and IV are used. Choice for feature 3 is [mild] unless:

$$F_{13,k}=0 \rightarrow F_{3,k}=[\text{none}].$$

$F_{13,k} > 1 \rightarrow$ the appropriate number of features of group B and C get the state [mild] or [moderate].

The choice for [severe] instead of [mild] will give identical results except for the presence of hyperplasia (feature 5).

In every run the matches between the simulated subject and all three targets (treatment classes) "normal", "suspect" or "positive" were calculated according to equation (4). The results for the eight runs are shown in Figures 5 and 6.

The model after adjusting the appropriate weighting factors shows the highest match percentage for the same target (class) as indicated by the tree (Figure 4) in all cases. The percentage for a diagnosis "positive" of a male animal (Figure 5) is 0% when no deviating feature is found, in contrast to a diagnosis of a female animal (Figure 6) where an elevated duct ratio can be found in combination with # deviating features = 0. For the same reason is the difference between the diagnoses "normal" and "positive" smaller for male animals (Figure 5d) than for female animals (Figure 6d) in the case that # deviating features = 1. In general, the comparable situations as illustrated in Figures 5a and 6a/b, in Figures 5c and 6c, and in Figures 5d and 6d respectively, shows highly comparable results. The addition of the

features of group III (Figures 5b, 5d, 6d) modifies the outcome of the model in the sense that in a lot of cases not 100% score can be reached. This reflects the situation that the finding of metaplasia (male) or the combination of metaplasia and an elevated duct ratio (female) accompanied with only a few or even no other deviations is unlikely or highly unlikely.

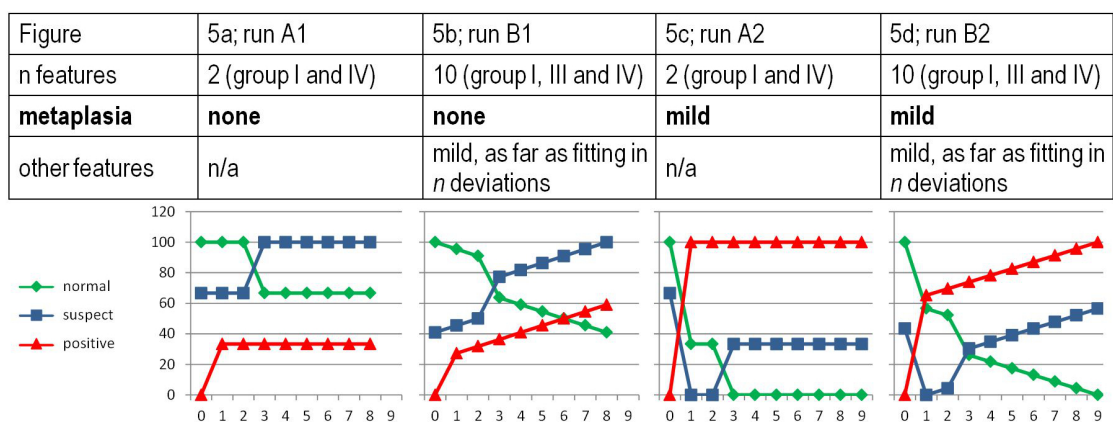


Figure 5. Performance of the free access key (matrix) for the prostate. Four different runs are illustrated. The choices for the main features are indicated in the tables on top of the figures. The choices for feature 13 (number of deviating features) running from zero to nine are given on the x-axis. The match percentage (according to equation 4) is given on the y-axis. The main differentiating feature per animal type is printed bold.

The large coverage of the targets indicated as “positive” (Table 3) is caused by the situation that the model is focusing on the correct diagnosis of possible treatment minimising the possibility of having false negative results. In both cases for male and female calves the final diagnosis is based on one feature (see Table 2 and Figure 4), whereas the states of the other features are overruled.

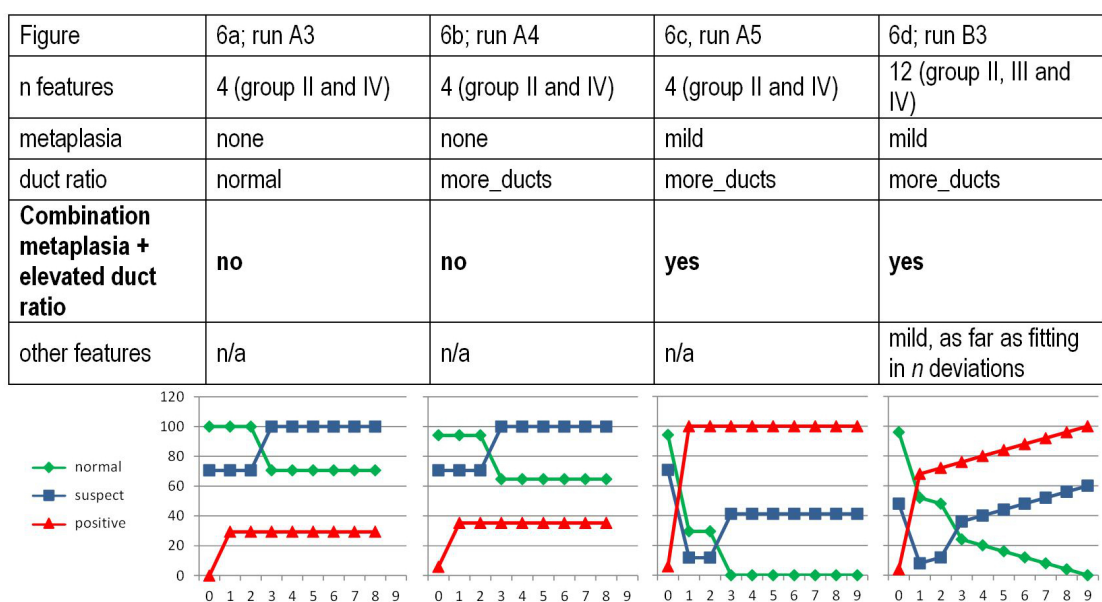


Figure 6. Performance of the free access key (matrix) for the gland of Bartholin. Four different runs are illustrated. The choices for the main features are indicated in the tables on top of the figures. The choices for feature 13 (number of deviating features) running from zero to nine are given on the x-axis. The match percentage (according to equation 4) is given on the y-axis. The main differentiating feature per animal type is printed bold.

Target	coverage	equation
001. Normal (prostate)	0.03	(11)
002. Suspect (prostate)	2.89	(11)
003. Positive (prostate)	34.68	(11)
004. Normal (Bartholins gland)	0.01	(11)
005. Suspect (Bartholins gland)	13.49	(11)
006. Positive (Bartholins gland)	34.68	(11)
Total coverage	85.79	(12)

Table 3. Coverage of the variability space by the individual targets and the total dataset.

The correlation between the features is shown in Table 4. Only a full correlation is found between the two features indicating the presence of metaplasia. This feature is included twice since different weighting factors appeared to be needed for the different animal types. Another reasonable high correlation factor was found between the duct ratio and the combined presence of metaplasia and elevated duct ratio. The presented level of correlation coefficients is in line with the calculated average redundancy: 0.405 (equation (8)).

	Feature	001. Metaplasia (male)	002. Duct ratio	003. Metaplasia (female)	004. Combinatio	005. # of features
▶	001. Metaplasia (male)	1.000	0.537	1.000	0.572	0.509
	002. Duct ratio	0.537	1.000	0.537	0.840	0.518
	003. Metaplasia (female)	1.000	0.537	1.000	0.572	0.509
	004. Combination	0.572	0.840	0.572	1.000	0.222
	005. # of features	0.509	0.518	0.509	0.222	1.000

Table 4. Matrix with Pearson's correlations between the features of the kernel model for diagnosis of illegal hormone use in veal calves. The colour of every cell (running from red to green) represents the value of the correlation coefficient.

The match table (Table 5) shows the relative resemblance between the targets based on equation (7). Except for the diagonal, the green colour, based on the calculations using equation (9), indicates that every target can be diagnosed uniquely compared to any other target. Hence, the separation capability is 100% (equation (10)).

	Target	001. Normal (prostate)	002. Suspect (prostate)	003. Positive (prostate)	004. Normal (Bartholir gland)	005. Suspect (Bartholir gland)	006. Positive (Bartholir gland)
▶	001. Normal (prostate)	1.000	0.733	0.373	0.667	0.333	0.173
	002. Suspect (prostate)	0.733	1.000	0.533	0.400	0.438	0.333
	003. Positive (prostate)	0.373	0.533	1.000	0.040	0.356	0.800
	004. Normal (Bartholins gland)	0.667	0.400	0.040	1.000	0.500	0.040
	005. Suspect (Bartholins gland)	0.333	0.438	0.356	0.500	1.000	0.356
	006. Positive (Bartholins gland)	0.173	0.333	0.800	0.040	0.356	1.000

Table 5. Matrix with the matches between the targets of the model for diagnosis of illegal hormone use in veal calves. The figure in every cell is calculated according to equation (7), the colour of every cell is based on equation (9).

4. Discussion

The process of identifying the level of treatment with growth hormones of veal calves is a rather specific situation for diagnosing in the broader framework of application of DSS in medicine [8-10]. Only one feature matters, all other features will only modify the probability that a diagnosis belongs to the correct class. Besides that, a constraint dependency rule exists between feature 13 (number of deviating features; Table 1) and the total of features from group III plus either from group I or group II which show a state other than normal. The importance of the main features is visible in Table 2 and Figure 4. The two main features (male: presence of metaplasia, female: combined presence of metaplasia and an elevated duct ratio) both got a weighting factor of 9 in order to outnumber the features in group III for reaching a correct diagnosis (number of features in group III plus 1). Since the presence of metaplasia in the diagnosis of a female calf does not form the exclusive indicator for treatment in contrast to the position of that feature in the diagnosis of the male calf, it got a weighting factor of only 1. The weight factors in the current model are fixed instead of being input sensitive [11].

There is no generic method for validation of data models in expert systems [7]. In the current study a top down modelling approach was chosen: logic tables lead to a decision tree, which was the basis for the full matrix of the free access key. This approach does not provide a tool for handling constraint dependency rules [7], which was solved here by optimising the weighting factors. Rass et al. [12] listed a number of requirements for valid expert systems. Of these, the requirements for minimising the redundancy and for avoiding unintended synonyms are now supported by measures to calculate the extent of these parameters: redundancy (equation (8)) and separation capability (equation (10)), respectively.

The position of the features of group III (Table 1: indicating the individual deviating characteristics) in an extended diagnosis (Figures 4b, 4d, 5d) can be discussed in terms of fuzzy logic principles. In several experiments with fuzzy logic comparable results have been found [9, 13]. Here, probability or uncertainty is the basic aspect causing patterns in the model outcomes that can be explained as membership functions [13]. As an example, the presence of metaplasia in a prostate is a definite diagnosis for treatment with growth hormones ($n = 1$ in Figure 5c in concordance with the tree in Figure 4), but it is highly unlikely that with such a diagnosis none of the other features of group III (Table 1) would show a state deviating from normal. The probability that an animal with the sole presence of metaplasia belongs to membership class "positive" is only slightly higher than its membership to the class "suspect" ($n=1$ in Figure 5d). The kernel model without using the individual features of group III (strategy A) seems sufficient to reach a diagnosis. All the features underlying the depending feature 13 (group IV) are nevertheless included in the model in order to improve the performance of the user by supporting his or her examinations, and to provide the possibility of an iterative process of optimising the diagnosis [14].

Existing results of optimising a data model for reaching a diagnosis reveal that lower numbers of features appeared to be optimal [10]. In those cases that a model consists of only a few features, expressing them in terms of space dimensions (e.g. a two-dimensional space in

Figure 2), a major part of the variation space might be covered. Increasing numbers of features (i.e. dimensions) result in an exponentially growing number of theoretically existing feature combinations that are not linked to a target. In the present study a total of approx. 14 % of the variability space was not covered by any target (Table 3). In order to evaluate this non-assigned part of the variability space, let us assume a variable number of features n each consisting of three feature states, a number of targets that can be accommodated by increasing with a factor of 2 with every additional feature, and one and only one state per feature identifying a target p :

$$(F_{i,k} \Rightarrow T_p) \wedge (F_{i,\neg k} \not\Rightarrow T_p) \quad (13)$$

Whereas equations (1) and (2) apply.

The resulting multidimensional spaces for a number of features ranging from 2 to 8, the number of targets accommodated and the resulting coverage are shown in Table 6. If more than one state of a feature can identify a target a larger coverage can be expected. This is the case in the here presented datamodel for the diagnosis of hormone treatment, since the probability to correctly classify all situations of hormone treatment was maximised. This is illustrated in Table 3. The high coverage of approx. 85.8% of the current model can be explained by the situation that the model was optimised to find all occasions of illegal use of hormones, i.e. the coverage of the classes “positive” was maximised.

n	Combinations (3^n)	Targets (2^n)	coverage
2	9	4	44%
3	27	8	30%
4	81	16	20%
5	243	32	13%
6	729	64	9%
7	2187	128	6%
8	6561	256	4%

Table 6. Relationship between the number of dimensions of a variability space (n), the possible number of combinations of feature states, and the coverage of the associated number of targets under the assumption of only one state per feature identifying a target (equation (13)).

The development of a specific model for reaching a histological diagnosis in a general platform provides several constraints, such as the lack of automatically calculating the number of deviating features (feature 13) from the number of individually selected features of group III. The advantage of the current procedure is the strict framework which forces to analyse the information structure in detail, and generic tools are available for testing and evaluation.

5. Conclusion

The presented parameters for redundancy, uniqueness, separation capability and coverage of variability space provide useful tools for the validation of a datamodel. The Developer as part of the Determinator system implements these parameters in an ordered manner, as exemplified in Table 5. The development and performance of the datamodel for reaching a diagnosis of the treatment of veal calves with hormones in the framework of Determinator reveals that a specific model can be developed and applied successfully in a generic framework.

Author details

L.W.D. van Raamsdonk^{1*}, S. van der Vange¹, M. Uiterwijk² and M. J. Groot¹

*Address all correspondence to: Leo.vanraamsdonk@wur.nl

1 RIKILT, Wageningen UR, Wageningen, the Netherlands

2 Alterra, Wageningen UR, Wageningen, the Netherlands

References

- [1] Raamsdonk, L. W. D., Mulder, P., & Uiterwijk, M. (2010). Identification tools as part of Feedsafety research: the case of ragwort. In: P.L. Nimis and R. VignesLebbe, *Proceedings of Bioidentify.eu: "Tools for identifying biodiversity: progress and problems"*, Edizioni Università di Trieste, 213-216.
- [2] Uiterwijk, M., van Raamsdonk, L. W. D., & Janssen, S. J. C. (2012). Determinator- a generic DSS for hazard identification of species or other physical subjects. In: R. Sepelt, A.A. Voinov, S. Lange, D. Bankamp (eds.), *Managing resources of a limited planet, sixth international congress of the International Environmental Modelling and Software Society. Leipzig, July 1-5.*
- [3] Hagedorn, G., Rambold, G., & Martellos, S. (2010). Types of identification keys. In: P.L. Nimis and R. VignesLebbe, *Proceedings of Bioidentify.eu: "Tools for identifying biodiversity: progress and problems"*, Edizioni Università di Trieste, 59-64.
- [4] Yost, R., Attanandana, T., Pierce Colfer, C. J., & Itoga, S. (2011). Decision Support Systems in agriculture: some successes and a bright future. In: C.S. Jao (ed.), *Efficient decision support systems: practice and challenges from current to future*, 291-330, Intech, Rijeka, Croatia.
- [5] Groot, M. J., Ossenkoppele, J. S., Bakker, R., Pfaffl, M. W., Meyer, H. H., & Nielen, M. W. (2007). Reference histology of veal calf genital and endocrine tissues- an update for screening on hormonal growth promoters. *J. Vet. Med. A Physiol Pathol Clin. Med.*, 54, 238-246.

- [6] Groot, M. J., Schilt, R., Ossenkoppele, J. S., Berende, P. L. M., & Haasnoot, W. (1998). Combinations of growth promoters in veal calves: consequences for screening and confirmation methods. *J. Vet. Med. A*, 45, 425-440.
- [7] Elfaki, A. O., Muthaiyah, S., Ho, C. K., & Phon-Amnuaisuk, S. (2011). Knowledge representation and validation in a Decision Support System: Introducing a variability modelling technique. In: C.S. Jao (ed.), *Efficient decision support systems: practice and challenges from current to future*, 29-48, Intech, Rijeka, Croatia.
- [8] Worachartcheewana, A., Nantasenamata, C., Isarankura-Na-Ayudhya, C., Pidetchab, P., & Prachayasittikula, V. (2010). Identification of metabolic syndrome using decision tree analysis. *Diabetes Research and Clinical Practice*, 90, e15-e18.
- [9] Das, A., & Bhattacharya, M. (2011). Computerized Decision Support System for Mass Identification in Breast Using Digital Mammogram: A Study on GA-Based Neuro-Fuzzy Approaches. In: H.R. Arabnia and Q.-N. Tran, *Software tools and algorithms for biological systems, Advances in Experimental Medicine and Biology* Springer Science and Business Media.
- [10] Sidiropoulos, K., Glotsos, D., Kostopoulos, S., Ravazoula, P., Kalatzis, I., Cavouros, D., & Stonham, J. (2012). Real time decision support system for diagnosis of rare cancers, trained in parallel, on a graphics processing unit. *Computers in Biology and Medicine*, 42, 376-386.
- [11] Kacprzyk, J., & Zadrozny, S. (2009). Supporting decision making via verbalization of data analysis using linguistic data summaries. In: E. Rakus-Andersson et al., *Recent Advances in Decision Making*, 45-66, Springer Verlag, Berlin, Heidelberg.
- [12] Rass, S., Machot, F. A., & Kyamakya, K. (2011). Fine-grained diagnostics of ontologies with assurance. In: C.S. Jao (ed.), *Efficient decision support systems: practice and challenges from current to future*, 79-98, Intech, Rijeka, Croatia.
- [13] Chen, G., Yan, P., & Wei, Q. (2009). Discovering associations with uncertainty from large databases. In: E. Rakus-Andersson et al., *Recent Advances in Decision Making*, 45-66, Springer Verlag, Berlin, Heidelberg.
- [14] Barthelemy, J. P., Bisdorff, R., & Coppin, G. (2002). Human centered processes and decision support systems. *European Journal of Operational Research*, 136, 233-252.

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen