Evaluating Tests in Medical Diagnosis: Combining Machine Learning with Game-Theoretical Concepts

Karlson Pfannschmidt[1], Eyke Hüllermeier[1], Susanne Held[2] and Reto Neiger[2]

[1] Department of Computer Science
Paderborn University, Germany
[2] Small Animal Clinic
Justus-Liebig University Gießen, Germany

Abstract. In medical diagnosis, information about the health state of a patient can often be obtained through different tests, which may perhaps be combined into an overall decision rule. Practically, this leads to several important questions. For example, which test or which subset of tests should be selected, taking into account the effectiveness of individual tests, synergies and redundancies between them, as well as their cost. How to produce an optimal decision rule on the basis of the data given, which typically consists of test results for patients with or without confirmed health condition. To address questions of this kind, we develop an approach that combines (semi-supervised) machine learning methodology with concepts from (cooperative) game theory. Roughly speaking, while the former is responsible for optimally combining single tests into decision rules, the latter is used to judge the influence and importance of individual tests as well as the interaction between them. Our approach is motivated and illustrated by a concrete case study in veterinary medicine, namely the diagnosis of a disease in cats called *feline infectious peritonitis*.

1 Introduction

Different types of tests, such as measuring serum antibody concentrations, are commonly used in medical diagnostics in order to reveal the health condition of an individual. The effectiveness of a single test is typically determined by correlating the test outcome with the true condition. Moreover, classical statistical hypothesis testing can be used to compare different test procedures in terms of their effectiveness.

In this paper, we tackle the problem of evaluating or selecting a test procedure from a slightly different perspective using methods of (semi-)supervised machine learning. Roughly speaking, the idea is that, by learning a model in which various candidate tests play the role of predictor variables, information about the usefulness of individual tests as well as their combination is provided by properties of that model. An approach of that kind has at least two important advantages:

- First, it not only allows for judging the usefulness of single tests but also of *combined tests*, i.e., the combination of different tests into one overall (diagnostic) decision rule. Thus, it informs about possible synergies (as well as redundancies) between individual tests and the potential to improve diagnostic accuracy thanks to a suitable combination of these tests.
- Second, going beyond the standard setting of supervised learning, a machine learning approach suggests various ways of improving the selection of tests by taking advantage of additional sources of information. An important special case is the use of *semi-supervised* learning to exploit "unlabeled" data coming from individuals for which tests have been made but the true health condition is unknown. This situation is highly relevant in medical practice, because tests can often be conducted quite easily, whereas determining the true health condition is very difficult or expensive.

Our approach is motivated by a concrete case study in veterinary medicine, namely the diagnosis of a disease in cats called *feline infectious peritonitis* (FIP). Complete certainty about whether or not a cat is FIP-positive, and eventually will die from the disease, requires a necropsy [10, 1]; unfortunately, no test performed in a cat while still alive has a 100% sensitivity or 100% specificity. Consequently, while different tests can be applied to cats quite easily, "labeling" a cat in the sense of supervised learning is expensive, difficult and time-consuming.

In addition to the use of (semi-supervised) machine learning methodology in medical diagnosis, we propose a game-theoretical approach for measuring the usefulness of individual tests as well as model-based combinations of such tests. Roughly speaking, the idea is to consider a combination of tests as a "coalition" in the sense of cooperative game theory, and the "payoff" of the coalition as the diagnostic accuracy achieved by the test combination. This approach will be detailed in the next section, prior to elaborating more closely on our case study in Section 3, presenting experimental results in Section 4 and concluding the paper in Section 5.

2 Evaluating single and combined tests

Suppose a set of tests $X_1, \ldots, X_K$ to be available. We consider the outcome of each test as a random variable $X_k : \Omega \longrightarrow \mathbb{R}$, where Ω is the population of individuals to which the test can be applied. Jointly, the K tests thus define a random vector

$$X = (X_1, \ldots, X_K) \in \mathcal{X} = \mathbb{R}^K \enspace .$$

The health state is a dichotomous variable $Y \in \mathcal{Y} = \{-1, +1\}$. Typically, each test is a positive indicator in the sense that $\mathbf{P}(Y = +1 \,|\, X_k)$ increases with X_k, i.e., the larger X_k, the larger the probability of the positive class. Using machine learning terminology, each test corresponds to a *feature* or predictor variable. Moreover, $\mathcal{X}$ is the *instance space*, each $X \in \mathcal{X}$ is an instance, and Y is the (binary) output or *response* variable.

2.1 Combined tests

If a diagnostic decision $\hat{y} \in \{-1, +1\}$ is not necessarily based on a single test X_k alone, but possibly uses a combination of several tests, a first question concerns the way in which such a combination is realized. From a machine learning point of view, this question is related to the choice of an underlying models class (hypothesis space)

$$\mathcal{H} \subset \bigcup_{J=1}^{K} \mathcal{H}_J = \bigcup_{J=1}^{K} \mathcal{Y}^{\mathbb{R}^J} \enspace ,$$

where $J \leq K$ is the number of tests included in the decision rule. Formally, we specify a combined test in terms of the subset $A \subseteq [K] = \{1, \ldots, K\}$ of indices, i.e., test X_k is included if $k \in A$.

The model class $\mathcal{H}$ could be defined, for example, as the class of linear threshold functions of the form

$$h : \big(x_{\sigma(1)}, \ldots, x_{\sigma(J)}\big) \mapsto \left[\!\!\left[\sum_{j=1}^{J} w_j \cdot x_{\sigma(j)} > t \right]\!\!\right] \enspace , \tag{1}$$

where $w_1, \ldots, w_J, t \in \mathbb{R}_+$ and $[\![\cdot]\!]$ maps true predicates to $+1$ and false predicates to -1; moreover, $\sigma(j)$ is the j-th test included in the combination, i.e., $\sigma(j) = k$ if $\sum_{i=1}^{k} [\![i \in A]\!] = j$.

2.2 Optimal decision rules

Let $L : \{-1, +1\}^2 \longrightarrow \mathbb{R}$ be a loss function, such that $L(y, \hat{y})$ denotes the penalty for making the diagnostic decision $\hat{y}$ if the true health state is y. For each combined test, specified by a subset $A \subseteq [K]$, there is an optimal decision rule

$$h_A^* \in \arg\min_{h \in \mathcal{H}} \int L\big(y, h(\boldsymbol{x})\big) \, d\,\mathbf{P}(\boldsymbol{x}, y) \enspace ,$$

i.e., a decision rule that minimizes the loss in expectation. We denote the expected loss of this model, which corresponds to the Bayes predictors in $\mathcal{H}_{|A|}$, by

$$e^*(A) = \int L\big(y, h_A^*(\boldsymbol{x})\big) \, d\,\mathbf{P}(\boldsymbol{x}, y) \enspace . \tag{2}$$

2.3 Estimating generalization performance

In practice, of course, neither the Bayes predictor h_A^* nor the ideal generalization performance $e^*(A)$ are known. Instead, we only assume a data set $\mathcal{D} = \mathcal{D}_L \cup \mathcal{D}_U$ to be given, which consists of a set of labeled instances

$$\mathcal{D}_L = \big\{(\boldsymbol{x}_i, y_i)\big\}_{i=1}^{L} \subset \mathcal{X} \times \mathcal{Y}$$

and possibly another set of unlabeled instances (test results without ground truth) $\mathcal{D}_U = \{\boldsymbol{x}_j\}_{j=1}^U \subset \mathcal{X}$. From a machine learning point of view, it is then natural to estimate the generalization performance on the basis of $\mathcal{D}$ for each $A \subseteq [K]$. To this end, models (1) can be fitted and their generalization performance can be estimated, for example, using cross-validation techniques or the bootstrap. More specifically, what can be estimated in this way is the generalization performance of a model that is trained on a combination A and data in the form of L labeled and U unlabeled examples. Therefore, we shall denote a corresponding estimate by $\hat{e}(A, L, U)$ or simply $\hat{e}(A)$ (assuming the underlying data to be given).

Needless to say, the estimates $\hat{e}(A)$ thus obtained are not necessarily monotone in the sense that $\hat{e}(B) \leq \hat{e}(A)$ for $A \subseteq B$. In fact, while $e^*(A)$ is the generalization performance of the Bayes predictor, i.e., the model that is obtained in the limit of an infinite sample size (provided the underlying learner is consistent), the estimates $\hat{e}(A)$ are obtained from models trained on a finite (and possibly small) data set. Therefore, practical problems such as overfitting become an issue, i.e., including additional tests may deteriorate instead of improve generalization performance.

2.4 Correcting generalization performance

How can the ideal generalization performances

$$\{e^*(A) \,|\, A \in [K]\} \tag{3}$$

be estimated? Starting with the finite-sample estimates

$$\big\{\hat{e}(A) \,|\, A \subseteq [K]\big\} \ , \tag{4}$$

our proposal is to correct these estimates so as to assure monotonicity. In fact, monotonicity is the main difference between the ideal and finite-sample scores. Apart from that, the ideal scores (3) should not differ too much from the estimates (4), i.e., $e^*(A) \approx \hat{e}(A)$, at least if the training data is not too small.

These considerations suggest the following estimation principle: Find a set of values (3) that satisfy monotonicity while remaining as close as possible to the corresponding scores (4). This principle can be formalized as an optimization problem of the following kind:

$$\begin{aligned} \text{minimize} \quad & \sum_{A \subseteq [K]} \big|\hat{e}(A) - e^*(A)\big| \\ \text{s.t.} \quad & \\ & e^*(B) \leq e^*(A) \text{ for all } A \subseteq B \subseteq [K] \\ & 0 \leq e^*(A) \leq 1 \text{ for all } A \subseteq [K] \end{aligned}$$

The above problem can be tackled by means of methods for *isotonic regression*. More specifically, since the inclusion relation on subsets induces a partial order on $2^{[K]}$, methods for isotonic regression on partially ordered structures are needed [3, 14].

2.5 Measuring the usefulness of tests

Consider the set function $\nu' : 2^{[K]} \longrightarrow [0,1]$ defined by $\nu'(A) = 1 - e^*(A)$. Obviously, ν' is a monotone measure (of the usefulness of combined tests). Moreover, this measure can be normalized by setting

$$\nu^*(A) = \frac{\nu'(A) - \nu'(\emptyset)}{\nu'([K]) - \nu'(\emptyset)} \ ,$$

where $\nu'(\emptyset)$ is the performance of the best (default) decision rule that does not use any test, i.e., which either always predicts $\hat{y} = +1$ or always $\hat{y} = -1$. The measure $\nu^*(\cdot)$ thus defined satisfies the following properties:

- $\nu^*(\emptyset) = 0$, $\nu^*([K]) = 1$,
- $\nu^*(A) \leq \nu^*(B)$ for all $A \subseteq B \subseteq [K]$.

Thus, ν^* is a normalized, monotone (but not necessarily additive) set function, referred to as *fuzzy measure* or *capacity* in the literature [5]. For each combined test A, $\nu^*(A)$ is a reasonable measure of the usefulness of this test.

In a similar way, a measure $v^\bullet$ can be defined on the basis of the finite-sample scores (4), that is, by normalizing $\nu'(A) = 1 - \hat{e}(A)$:

$$v^\bullet(A) = \frac{\nu'(A) - \nu'_{min}}{\nu'_{max} - \nu'_{min}} \ ,$$

where $\nu'_{min} = 1 - \max_{B \subseteq [K]} \hat{e}(B)$ and $\nu'_{max} = 1 - \min_{B \subseteq [K]} \hat{e}(B)$. Note, however, that this measure is not necessarily monotone.

Which of the two measures is more meaningful, ν^* or $\nu^\bullet$? The answer to this question depends on practical considerations and what the measure is actually supposed to capture. When being interested in the *potential* asymptotic usefulness of a test combination, then ν^* is the right measure. Otherwise, if a model induced from a concrete set of training data is supposed to be put into (medical) practice, $\nu^\bullet$ is arguably more relevant.

2.6 Shapley value and interaction index

From the point of view of (cooperative) game theory, each (test) combination $A \subseteq [K]$ can be seen as a *coalition* and $\nu \in \{\nu^*, \nu^\bullet\}$ as the *characteristic function*, i.e., $v(A)$ is the *payoff* achieved by the coalition A. Thanks to this view, we can take advantage of various established game-theoretical concepts for analyzing the importance of individual players, which correspond to tests in our case, as well as the interaction between them. In particular, the *Shapley value*, also called importance index, is defined as follows [17]:

$$\varphi(k) = \sum_{A \subseteq [K] \setminus \{k\}} \frac{1}{K \binom{K-1}{|A|}} \Big(\nu(A \cup \{k\}) - \nu(A) \Big) \ . \tag{5}$$

The Shapley value of ν is the vector $\boldsymbol{\varphi}(\nu) = (\varphi(1), \ldots, \varphi(K))$. For monotone measures (such as $\nu = \nu^*$), one can show that $0 \leq \varphi(k) \leq 1$ and $\sum_{k=1}^{K} \varphi(k) = 1$; thus, $\varphi(k)$ is a measure of the *relative* importance of the test X_k.

The *interaction index*, as proposed by [13], is defined as follows:

$$I(i,j) = \sum_{A \subseteq [K] \setminus \{i,j\}} \frac{\Big(\nu(A \cup \{i,j\}) - \nu(A \cup \{i\}) - \nu(A \cup \{j\}) + \nu(A)\Big)}{(K-1)\binom{K-2}{|A|}} .$$

This index ranges between -1 and $+1$ and indicates a positive (negative) interaction between the tests X_i and X_j if $I_{i,j} > 0$ ($I_{i,j} < 0$).

It is worth mentioning that the approach put forward in this section is quite in line with the idea of *Shapley value regression* [11], which makes use of the Shapley value in order to quantify the contribution of predictor variables in (linear) regression analysis (quantifying the value of a set of variables in terms of the R^2 measure on the training data).

3 Feline infectious peritonitis in cats

Feline infectious peritonitis (FIP) is a disease with an affinity to young cats, a predisposition to involve cats living in larger groups. As it exhibits typical physical examination and clinical laboratory findings, it appears to be easy to diagnose. However, while a presumptive diagnosis is quickly established, a definite diagnosis is difficult to impossible to obtain without gross and histopathological evaluation including immunohistochemistry [10, 1].

The seroprevalence is high, especially in catteries where up to 90 % of the cats are positive [2], but also up to 50 % of cats living in single-cat households have coronavirus-specific antibodies [4]. Of these, 5-10 % will develop the deadly form of FIP. A characteristic symptom of FIP is body cavity effusion, which also appears in other diseases [8]. Several treatment options exist for some of these diseases while FIP is deadly and no reliable effective therapy is known so far [16]. Therefore, it is important to diagnose the correct disease early.

Several diagnostic tests are available that diagnose FIP, for which sensitivity, specificity, positive and negative predictive value vary between different studies, presumably because different forms of FIP (effusive and dry) were investigated and because various clinical signs, geographic locations, years of investigation, prevalence and combination of tests were used [15, 7, 6, 4, 9, 18]. In studies so far, no cat had all available tests performed.

The data underlying our study includes the following diagnostic tests:

- Albumin to Globulin ratio, plasma (X_1) and effusion (X_2)
- Rivalta test (X_3)
- Presence of antibodies against feline coronavirus (FCoV, X_4)
- Reverse transcriptase nested polymerase chain reaction (RT-nPCR) to detect FCoV-RNA in EDTA-blood (X_5) and in the effusion (X_6)
- Immunofluorescence staining (IFA) of FCoV antigen in macrophages in the effusion (X_7)

4 Empirical study

Our dataset consists of 100 cats in total. For 29 of these cats, a necropsy was performed to establish the gold standard diagnosis; 11 of the 29 cats were diagnosed with feline infectious peritonitis (FIP). Additionally, the above 7 diagnostic tests were performed on all cats (i.e., $K = 7$, $L = 29$ and $U = 71$).

To estimate the generalization accuracy (in terms of the simple 0/1 loss function) of each of the $2^7 = 128$ combined diagnostic tests, we employ a semi-supervised classification technique called maximum contrastive pessimistic likelihood estimation (MCPL) [12]. Logistic regression with L_2 penalization is used as the base learner in MCPL, i.e., individual tests are combined using a linear model of the form (1).

Estimates $\hat{e}(A)$ of the (finite-sample) classification errors are obtained as follows: We resample the set of 29 labelled cats and split the resulting sample into 16 training and 13 test examples. The remaining 71 cats without label information are added to the training set. This procedure is repeated 501 times for each of the 128 combinations of tests, and the results are averaged. To obtain estimates $e^*(A)$ of the ideal generalization performances, the finite-sample estimates are subsequently corrected using isotonic regression [3, 14] as described in Section 2.4.

4.1 Test importance for finite-sample performance estimates

Figure 1 shows the Shapley values calculated for each test on the basis of the finite-sample performances $\hat{e}(A)$, i.e., the measure $\nu^\bullet$. Note that, since this measure is not necessarily monotone, negative Shapley values are possible (as is the case for the Rivalta test). The highest Shapley values are obtained for the two RT-nPCR tests.

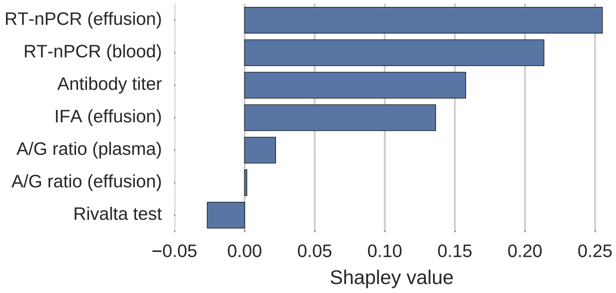


Fig. 1. Shapley values calculated for the finite-sample measure $\nu^\bullet$.

To further illustrate the importance of the diagnostic test RT-nPCR, Figure 2 shows the mean validated classification accuracy for all 128 test combinations. The 80 % empirical percentiles are indicated by the vertical lines, and the subsets

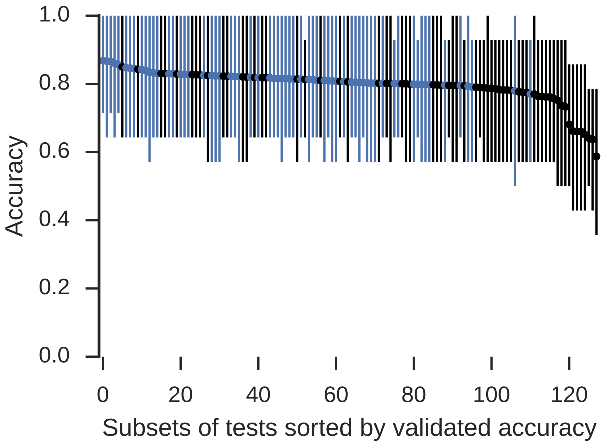


Fig. 2. Classification accuracy for all 128 test combinations (sorted by mean accuracy). The vertical lines show the 80 % empirical percentiles of the bootstrap estimates. The results for subsets including RT-nPCR (blood) are highlighted in blue.

are sorted in decreasing order of their mean validated accuracy. Moreover, the results for those subsets including RT-nPCR (measured in blood) are highlighted in blue. Evidently, the concentration of subsets containing RT-nPCR (blood) is systematically higher to the left of the plot, which confirms that the inclusion of the test improves diagnostic accuracy.

4.2 Test importance for ideal performance estimates

The effect of isotonic regression on the finite-sample estimates is shown in Figure 3. Here, each blue dot corresponds to an estimate $\hat{e}(A)$ for a particular subset A of diagnostic tests. Since partial monotonicity, which is assured by isotonic regression, cannot be visualized in a two-dimensional plot, the data points are sorted by their corrected classification accuracy (and ties are broken at random). The green line shows the isotonic regression fit.

The corrected performance estimates $\nu^*(A)$ can subsequently be used to calculate the Shapley values for each diagnostic test. The results are shown in Figure 4. Due to the monotonicity of ν^*, all values are now positive. Again, the RT-nPCR tests achieve the highest Shapley values, but FCoV antibody titer and IFA (effusion) obtain values > 0.15, too. Note that the relative order of the RT-nPCR tests changed from the one in Figure 1, probably due to their accuracy being very similar and the random nature of the bootstrap validation.

Figure 5 shows the accuracy estimates for all subsets. The dots indicate the corrected accuracies $\nu^*(A)$ and are used to sort subsets in decreasing order, while the vertical lines show the 80 % percentiles of the original bootstrap estimates. Again, the results are highlighted in blue if RT-nPCR (blood) is included in A.

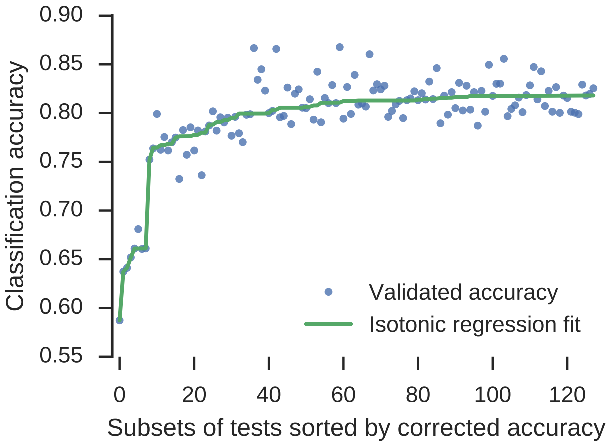


Fig. 3. Isotonic regression correction (green line) applied to the bootstrap validated classification accuracies (blue dots).

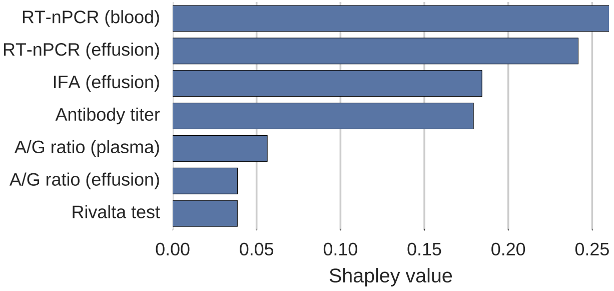


Fig. 4. Shapley values calculated using the corrected validation accuracies.

Like in the case of $\nu^\bullet$ (cf. Figure 2), the subsets containing RT-nPCR (blood) can mostly be found on the left side of the plot; this trend is now even more pronounced.

4.3 Balancing accuracy and cost

An important question for a veterinary physician is which combination A of tests to perform, taking into account both diagnostic accuracy and effort. Figure 6 shows the corrected accuracies $\nu^*(A)$ (green dots) of all subsets of tests and their combined monetary cost in Euro. The Pareto set, consisting of those combinations that are not outperformed by any other combination in terms of both accuracy and cost at the same time, is indicated as a blue line. From a practical point of view, the result suggests to use a single diagnostic test, namely RT-nPCR (blood

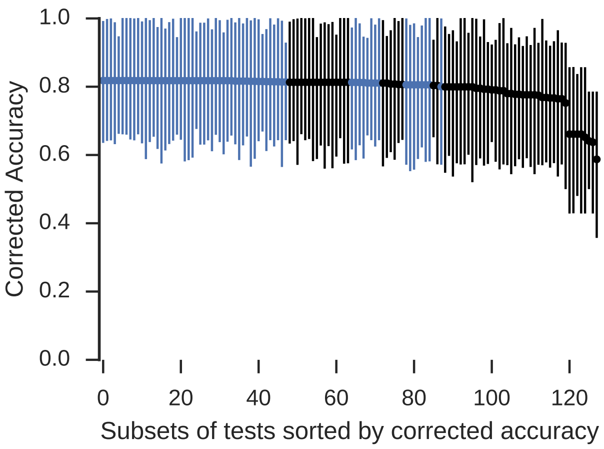


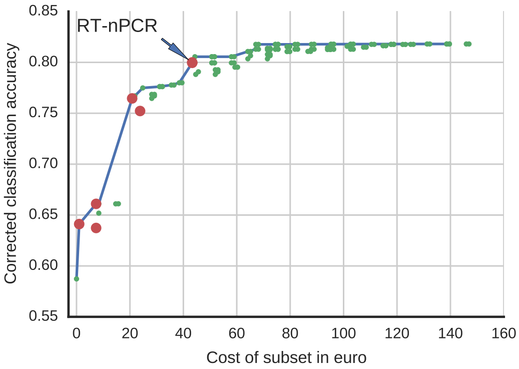
Fig. 5. Corrected accuracy $\nu^*(A)$ for all 128 subsets (sorted by mean accuracy). The vertical lines show the 80 % empirical percentiles of the original bootstrap estimates. Subsets including RT-nPCR (blood) are shown in blue.

or effusion), because the inclusion of more tests yields only minor improvements. This is confirmed by the pairwise interaction indices shown for both measures $\nu^\bullet$ and ν^* in Table 1. All these measures are negative, suggesting that the tests are more redundant than complementary.

Note that, once a decision in favor of using a single test is made, the Shapley value, as a measure of average improvement achieved by adding a test, is no longer the best indicator of the usefulness of a test. Instead, a selection should be made based on the tests' individual performance. With a validated accuracy of 87 %, RT-nPCR (effusion) appears to be the best choice in this regard.

Table 1. Pairwise interaction indices for $\nu^\bullet$ (left) and ν^* (right).

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	0.0	−0.02 −0.03	−0.05 −0.05	0.00 −0.02	−0.20 −0.21	−0.15 −0.21	−0.17 −0.19
X_2		0.0	−0.05 −0.05	−0.02 −0.03	−0.07 −0.05	−0.03 −0.04	−0.04 −0.04
X_3			0.0	−0.02 −0.04	−0.10 −0.06	−0.04 −0.06	−0.07 −0.06
X_4				0.0	−0.08 −0.05	−0.01 −0.04	−0.04 −0.03
X_5					0.0	−0.30 −0.33	−0.24 −0.32
X_6						0.0	−0.20 −0.22



Diagnostic test	Cost
A/G ratio (plasma)	€ 7.34
A/G ratio (effusion)	€ 7.34
Rivalta test	€ 1.00
Antibody titer	€20.83
RT-nPCR (blood)	€43.32
RT-nPCR (effusion)	€43.32
IFA (effusion)	€23.80

Fig. 6. Scatter plot of the monetary costs of the subsets in Euro in relation to the corrected accuracies $\nu^*(A)$ shown as green dots. The blue line shows the Pareto front. The red points highlight the subsets which contain exactly one test. The costs for each individual test are shown in the table on the right.

5 Summary and conclusion

In this paper, we proposed a method for measuring the importance and usefulness of predictor variables in (semi-/supervised) machine learning, which makes use of concepts from cooperative game theory: subsets of variables are considered as coalitions, and their predictive performance plays the role of the payoff. Although our approach is motivated by a concrete application in veterinary medicine, namely the diagnosis of *feline infectious peritonitis* in cats, it is completely general and can obviously be used for other learning problems as well.

For the case study just mentioned, our method produces results that appear to be plausible and agree with the medical experts' experience. Roughly speaking, there are two strong diagnostic tests that are significantly more accurate than others; practically, it suffices to use one of them, since a combination with other tests yields only minor improvements.

There are several directions for future work. For example, the principle we proposed in Section 2.4 for inducing ideal generalization performances $e^*(A)$ from finite-sample estimates $\hat{e}(A)$ is clearly plausible and, moreover, seems to be indeed able to calibrate the original estimates thanks to an ensemble effect. Nevertheless, it calls for a more thorough analysis and theoretical justification.

References

1. Addie, D.D., Paltrinieri, S., Pedersen, N.C.: Recommendations from workshops of the second international feline coronavirus/feline infectious peritonitis symposium. Journal of feline medicine and surgery 6(2), 125–130 (2004)

2. Benetka, V., Kübber-Heiss, A., Kolodziejek, J., Nowotny, N., Hofmann-Parisot, M., Möstl, K.: Prevalence of feline coronavirus types I and II in cats with histopathologically verified feline infectious peritonitis. Veterinary microbiology 99(1), 31–42 (2004)
3. Block, H., Qian, S., Sampson, A.: Structure algorithms for partially ordered isotonic regression. Journal of Computational and Graphical Statistics 3(3), 285–300 (1994)
4. Giori, L., Giordano, A., Giudice, C., Grieco, V., Paltrinieri, S.: Performances of different diagnostic tests for feline infectious peritonitis in challenging clinical cases. Journal of Small Animal Practice 52(3), 152–157 (2011)
5. Grabisch, M., Nguyen, H., Walker, E.: Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference. Kluwer Academic Publishers (1995)
6. Hartmann, K., Binder, C., Hirschberger, J., Cole, D., Reinacher, M., Schroo, S., Frost, J., Egberink, H., Lutz, H., Hermanns, W.: Comparison of Different Tests to Diagnose Feline Infectious Peritonitis. Journal of Veterinary Internal Medicine 17(6), 781–790 (2003)
7. Hirschberger, J., Hartmann, K., Wilhelm, N., Frost, J., Kraft, W.: Using direct immunofluorescence to detect coronaviruses in peritoneal in peritoneal and pleural effusions. Tierärztliche Praxis 23, 92–99 (1995)
8. Hirschberger, J., DeNicola, D.B., Hermanns, W., Kraft, W.: Sensitivity and specificity of cytologic evaluation in the diagnosis of neoplasia in body fluids from dogs and cats. Veterinary Clinical Pathology 28(4), 142–146 (1999)
9. Jeffery, U., Deitz, K., Hostetter, S.: Positive predictive value of albumin: globulin ratio for feline infectious peritonitis in a mid-western referral hospital population. Journal of feline medicine and surgery 14(12), 903–905 (2012)
10. Kipar, A., Köhler, K., Leukert, W., Reinacher, M.: A comparison of lymphatic tissues from cats with spontaneous feline infectious peritonitis (FIP), cats with FIP virus infection but no FIP, and cats with no infection. Journal of comparative pathology 125(2), 182–191 (2001)
11. Lipovetsky, S., Conklin, M.: Analysis of regression in game theory approach. Applied Stochastic Models in Business and Industry 17(4), 319–330 (2001)
12. Loog, M.: Contrastive pessimistic likelihood estimation for semi-supervised classification. IEEE Transactions on Pattern Analysis and Machine Intelligence PP(99) (2015)
13. Murofushi, T., Soneda, S.: Techniques for reading fuzzy measures (iii): interaction index. In: 9th Fuzzy System Symposium, pp. 693–696. Sapporo, Japan (1993)
14. Pardalos, P., Xue, G.: Algorithms for a class of isotonic regression problems. Algorithmica 23(3), 211–222 (1999)
15. Parodi, M.C., Cammarata, G., Paltrinieri, S., Lavazza, A., Ape, F.: Using direct immunofluorescence to detect coronaviruses in peritoneal and pleural effusions. Journal of Small Animal Practice 34(12), 609–613 (1993)
16. Ritz, S., Egberink, H., Hartmann, K.: Effect of Feline Interferon-Omega on the Survival Time and Quality of Life of Cats with Feline Infectious Peritonitis. Journal of veterinary internal medicine 21(6), 1193–1197 (2007)
17. Shapley, L.: A value for n-person games. Annals of Mathematical Studies 28, 307–317 (1953)
18. Soma, T., Wada, M., Taharaguchi, S., Tajima, T.: Detection of ascitic feline coronavirus RNA from cats with clinically suspected feline infectious peritonitis. Journal of Veterinary Medical Science 75(10), 1389–1392 (2013)