THE APPLICATION OF ARTIFICIAL INTELLIGENCE FOR THE ANALYSIS OF DATA ACQUIRED BY IN-LINE INSPECTION OF PIPELINES

K. Reber[1], H. Willems[1], A.O. Barbian[1]; M. Zoellner[2], M. Ziegenmeyer[2]
[1]NDT Systems & Services AG, Stutensee, Germany, [2]Forschungszentrum Informatik, Group IDS, Karlsruhe, Germany

Abstract: With increasing automation of NDE-measurements and the implied necessity to quickly scan through large amounts of data for the localisation of flaw candidates, there is an increasing demand for computer-based know-how.
At NDT Systems & Services AG there is the demand to quickly analyse NDE-data that has been acquired by intelligent pigs used for in-line inspection of pipelines. Usually the problem is solved by finding defect candidates first using simple methods with minor discrimination power. In a second step a classification will distinguish between relevant flaws and irrelevant spurious signals. This is achieved by means of artificial intelligence.
In the past the classification task has often been carried out by neural networks. This contribution will demonstrate how the application of Support Vector Machines (SVM) has considerably improved the handling and later work with such a learning machine. A comparison with neural networks is given. The improvements with continued training are demonstrated. The topic of retraining, i.e. the extension of the knowledge is discussed with special respect to the selection of training samples and their archiving.

Introduction: For the reliability of non-destructive testing two aspects can be distinguished: the data acquisition technology and the signal evaluation. While improvements in the first field mainly consist of technological advancements, the latter has so far been an area, where the human factor plays an important role. Although it does not seem that the human input will soon be negligible, the introduction of artificial intelligence can at least introduce a basis of repeatability such that decision making in signal evaluation can be reviewed and is to some degree predictable.
The application of artificial intelligence is often viewed with the fear of rendering humans superfluous or at least less important and thus many arguments have been brought forward against it. They range from concerns about safety to the possibility of increased unemployment, the naming may actually contribute to this notion. If artificial intelligence is viewed more pragmatically, it should rather be considered an advanced helping hand and a standardization in decision making, where simple rule-based decisions are not possible.
Artificial intelligence has been introduced to non destructive testing for a long time now. Especially in automated processes it can contribute to achieve a higher degree of automation. In this paper a field shall be described where the amount of data to be checked for defects is so large that only the automation of signal evaluation can lead to data analysis times required. The inspection of pipelines is nowadays carried out with so-called intelligent pigs, inspection vehicles that travel through the pipeline with the product and carry out different kinds of inspection. We will focus on the ultrasonic inspection of pipelines for wall thickness reductions by metal loss and their data analysis. The intelligent pig acquires data self-sufficient for up to several days. The respective data is stored on-board. With the receiving to the tool a multitude of inspection data becomes available at once. Nevertheless a very quick and at the same time reliable assessment of the signals is required. As the operators of pipelines need to have a quick verification of the existence of severe defects, it has become common to deliver a preliminary report that states only the most critical defects. The full evaluation is delivered several weeks later. There is, however, a contradiction in finding all critical defects quickly without assessing all available signals. The criterion of what a "critical" defect really consists of is a matter of discussion. For most kinds of inspection critical defects do not necessarily show the largest signal amplitudes. If signals are not fully evaluated there is no proof that they do not belong to a critical defect.
The need for a tool to help in quickly evaluating large amounts of data is thus not relieved by the possibility to issue a preliminary report, but to the contrary such a tool can be applied already in the stage of generating a preliminary report.
The first solution for such a tool applying artificial intelligence can be found in [1],where neural networks are used. In a final stage features are categorized into different classes. It is typical for ultrasonic inspection data that the classification into different types of features is the final analysis step that is not straight forward. The size of the feature, i.e. in case of metal loss (like corrosion), is not a matter of interpretation in ultrasonic data. This is a

difference to MFL-inspection. Here an important aspect of data analysis consists of converting the signals into a defect geometry. For a signal not classified as metal loss no depth should be calculated.

The advantage of automatic feature classification: The automation of data analysis can be distinguished in several steps. If the situation is such that all tool data is available, the first step is to generate a pipe tally, i.e. a list of all girth welds. The second step consists of a search for defect candidates. For areas, where the wall thickness deviates from the usual local wall thickness, a box is drawn. In the language of data analysis this step is called boxing and really is the first automation step. Almost all operators of intelligent pigs use such an automation process. In the language of pattern recognition this is a segmentation of the data into irrelevant parts of the pipe and regions of interest.

The second step of automation is the classification step. In this step boxes are categorized into different classes. Naturally the value of this step highly depends on the quality of the first automation step, the boxing. Obviously there is an interaction between the two steps. The classification has to rely on a proper boxing of the defects to be able to make decisions. The boxing on the other side does not put its focus on a discrimination of irrelevant features as this is carried out later in the classification. The better the discrimination power of the boxing, the less important is the classification. Also if the density of features in the inspected pipeline (the actual degree of corrosion for instance) the less important the automation will be, because in this case most work would be carried out manually anyway.

For both steps errors occur. For the considerations below only the errors in the classification step shall be further addressed. A type one error would be an incorrect removal of a feature from the set of relevant features. This would lead to a missing of a defect and should in general be avoided. The larger the defect the less often type one errors should occur. Finally type two errors describe errors where a feature is considered relevant, although it is not. These errors are less dangerous, but still question the usability of the automation.

Advantages of Support Vector Machines: Support Vector Machines belong to the relatively new family of Kernel Methods, that combine the simplicity and computational efficiency of linear algorithms, such as the perceptron algorithm, with the flexibility of non-linear systems, such as for example neural networks, and the rigor of statistical approaches such as regularization methods in multivariate statistics [2]. By reducing the learning step to a convex optimization problem, which can always be solved in polynomial time, the problem of local minima typical of neural networks, decision trees and other non-linear approaches is avoided. Therefore the training of support vector machines is deterministic and retraining is faster and easier.

Moreover due to their foundation in the principles of Statistical Learning Theory they are remarkably resistant to overfitting especially in circumstances where other methods are affected by the ‘curse of dimensionality’.

The main idea of Kernel Methods is to first embed the data into a suitable vector space and then detect relevant patterns in the resulting set of points by using simple linear methods. If the embedding map is non-linear, one is enabled to discover non-linear relations by the means of linear algorithms. Such a mapping of itself will not solve the problem, but coupled with the following two observations it becomes very effective:

- The support vector algorithm only needs the information about the relative positions of the data points in the embedding space encoded in the inner products between them.
- The inner products between the projections of data inputs into high dimensional embedding spaces can be efficiently computed directly from the inputs using a so-called kernel function.

Support Vector classification provides a computationally efficient way of learning ‘good’ separating hyperplanes in a high dimensional vector space, where ‘good’ hyperplanes are ones that optimize generalization bounds, and ‘computationally efficient’ means algorithms that are able to deal with sample sizes of hundreds of thousands of examples. Due to the clear guidelines of generalization theory on how to control capacity, overfitting is prevented by controlling the hyperplane margin measures. The mathematical techniques necessary to find hyperplanes optimizing these measures are provided by optimization theory.

The absolute value of the resulting decision function is called activation. It corresponds to the distance between the projected data input and the separating hyperplane in the vector space. Therefore it indicates the classification quality for a given feature: a low activation stands for an unconfident classification and a high activation for a confident classification.

The actual implementation and performance tests: In the actual implementation feature classes have been set up for metal loss, laminations, dents, pipeline installations and inclusions. To complete the set there is also a class for features that cannot be clearly assigned to one of the classes, i.e. their class affiliation is ambiguous. Finally, of course, there is a class of irrelevant features, that do not show a material flaw but spurious signals. The method of how the class "ambiguous" is assigned is very important for the classification system in general. Naturally for all implementations of artificial intelligence there are ambiguous decisions, just like in real life. In the case of ultrasonic pipeline inspection there is often not a clear decision boundary on whether a feature is an inclusion in the pipe or a lamination. Laminations that are interrupted and very short can easily be considered inclusions. Similarly metal loss features that are very shallow do not necessarily represent corrosion and are rather irrelevant. A boundary has to be drawn that is arbitrary.

For SVM this ambiguity is manifested in different ways. If the learning has been incomplete no activations for any class are present. In this case as well as for activations that are below a specified threshold the class "ambiguous" is assigned. These assignments represent incomplete training, which is of course also possible for rather matured learning machines. If, to the contrary, the learning has advanced very far, multiple activations are possible. In such a case a rule can be conceived of that would tell what class activations overrules what other class activation. It was deemed clearer and for the beginning to also classify these features as ambiguous. In the analysis process features classified as "ambiguous" will always be checked again by data analysts. Moreover ambiguous features are not in the training set.

Performance Figures: To actually assess the results of benchmark tests with sequentially trained SVMs the figures of merit have to be clarified. Obviously the above mentioned type one and type two errors are such figures of merit. With the actual implementation other figures of merit are possible. To compare the results of the SVM with the correct classification a so-called confusion matrix can be set up. In this matrix the number of features that are classified into a certain class are compared to the number of features that really belong to this class. For the "real" class affiliation, a large set of data has to be manually classified by an expert, whose decisions are not to be questioned. The table below shows such a confusion matrix.

For instance this table shows, that the SVM has classified 32 features as laminations. Out of these 8 are really to be considered laminations and will have to appear in the inspection report. 24 are irrelevant features, that have to be discarded. All type two errors are bounded by a blue (dark gray) box. There are altogether 1007 type two errors (false calls) out of 15777 features to be classified. Type one errors are bounded by the red (light gray) box. There are altogether 7 errors of such a kind.

Other figures of merit are the number of features to be manually checked, the number of features classified correctly and the complement, the number of features classified incorrectly. The number of features to be checked is given by all features that are not classified as irrelevant, in this case 4742, or 30%. The features classified correctly in a sense that the SVM has produced the desired result are found on the diagonal of the matrix. For these features the training has been appropriate.

SVM / manual	irrelevant	installation	dent	inclusion	lamination	metal loss	ambiguous	sum
irrelevant	**11028**	**0**	**3**	**37**	**24**	**943**	**3438**	**15473**
installation	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
dent	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
inclusion	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**2**
lamination	**0**	**0**	**0**	**2**	**8**	**1**	**4**	**15**
metal loss	**6**	**0**	**1**	**0**	**0**	**267**	**13**	**287**

ambiguous	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
sum	**11035**	**0**	**4**	**39**	**32**	**1211**	**3456**	**15777**

Five learning machines have been trained with sets of increasing power. The previous training set is always a subset to the later training set. Training samples are taken from six different inspections. They have been selected to make the diversity as large as possible. The first set only consists of 1635 samples, the last one has 8273. With the advancement of training several aspects have to be checked.

Consistency with older training data: Features that have been correctly classified even with the smallest training set should not deteriorate too much with continued training. A set of features (204) has been selected from a specific inspection, that was correctly classified with the smallest training set. Many of such features can be found, because similar features are in the first training set. The correctness of the classification has been checked for all learning machines with larger training sets. The performance is to be tested with the figures of merit described above. The result is shown in Figure 1.

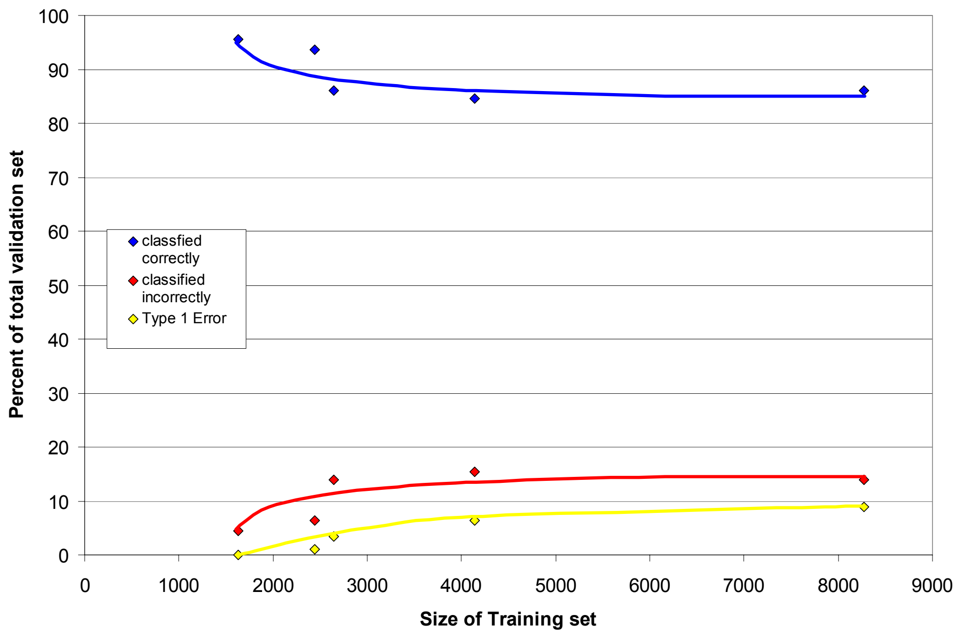

Figure 1: Change of Performance parameters with continued training for a set of data that is classified correctly with a small training set.

The behavior is as expected in a sense that initially the number of correctly classified features slightly decreases. A further increase is not possible, because of the initial selection. The decrease takes place with little further training. It levels off for larger training sets. It is important that further training does not continuously decrease the performance on previously correctly classified features. In this case further training would not always be beneficial and training samples would have to be checked before including them. However, the decrease of performance here is small and just reflects the small number of ambiguous features that is inevitable.

Advancement of training: A set of features is selected from an inspection that is new to the SVM, i.e. no samples of this inspection are in any training set. Naturally the performance on the first SVM is poor. With increased knowledge represented in the training set the performance should improve. In Figure 2 the number of type two errors and the number of features to be checked again are shown. Type two errors decrease, which means that less false calls are being made. In addition the number of features to be checked also decreases. This is especially the case after the first retraining. Less features are called ambiguous. The SVM derives a decision for an

increasing number of features. Both curves prove how the larger knowledge base improves the performance in these two figures of merit.
In Figure 3 two more figures of merit are plotted. The number of features with no activation has a similar behavior as the features "to be checked" in Figure 2. There is a sudden drop in the beginning and a further decrease with further training. Especially the sudden drop at the first retraining can be explained if a closer look is taken at the training sets. In the second training set there are samples of features that were found in an inspection very similar to the validation set. Both pipelines were build at the same time (about 40 years ago) are made from the same type of pipe and the inspection was carried out in the same product. Thus the level of echo-loss, noise and other influences on the signal are the same and the feature vector is in close vicinity to the validation set in the feature vector space.

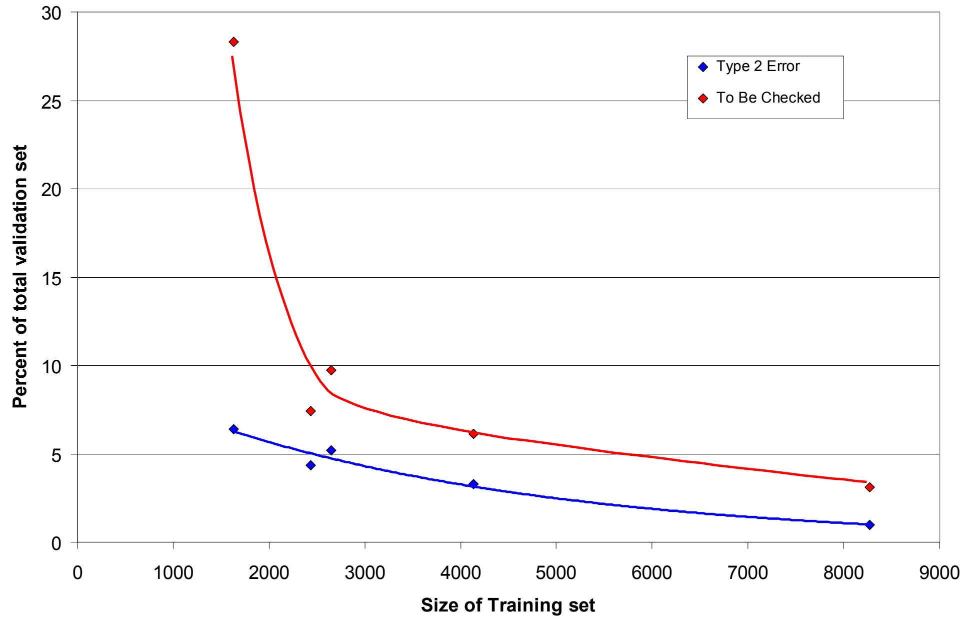

Figure 2: Performance figures for continued training on a set of data that is different from the training data.

The other curve in Figure 3 shows the change in number of errors of the first kind. This behavior was not anticipated. Also a drop is expected. Instead a steady increase is observed. Although the overall level of type two errors is small, it rises to about 1 percent, the behavior is not satisfactory and shall be further examined.

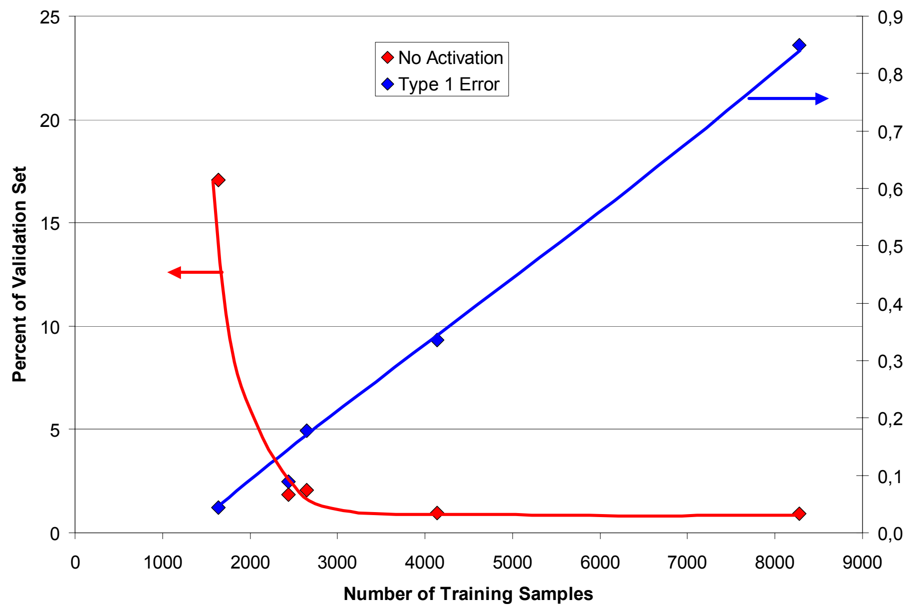

Figure 3: Two different performance figures for a data set with increasing performance.

A statistical analysis of all features that were incorrectly discarded was done. For the last training set the number of these features was 134. Two inclusions and three laminations were rejected which is in some cases, esp. for small sizes permissible. Also 129 metal loss objects were rejected. Naturally there has to be an arbitrary boundary of what kind of change in wall thickness is to be considered a metal loss and what is really irrelevant. In most cases an analysis threshold is agreed upon with the contractor (usually the pipeline operator). Typically metal loss type features have to be reported, if they exceed 1 mm decrease in remaining wall thickness. Figure 4 shows a depth histogram of these 129 features.

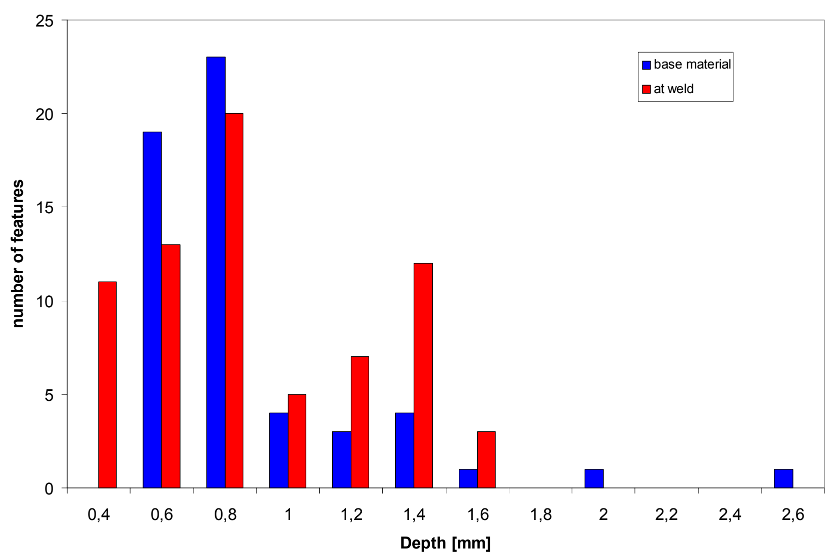

Figure 4: Depth histogram of features misclassified with a type 1 error. Most features are very shallow. The ambiguity of training samples is responsible for these type 1 errors.

Features are separated depending on whether they are located near a weld or in base material. Most of these metal loss type features especially in the base material are rather shallow. The fact that this type of error increases must be attributed to the increasing number of training samples originating from seamless pipeline. In this type of pipe there are many manufacturing related changes in wall thickness. A decrease in depth is thus less relevant as compared to welded pipe. In welded pipe it is much more likely to have an early corrosion site, which can be relevant. In this pipe type there are many training samples with remaining wall thickness reduction of only 0.6 mm. Even though this is hardly ever reported as metal loss in the inspection report, it was deemed that the decision on relevance in such a case should be left to the data analyst and not to the machine. In seamless pipe a loss of 0.6 mm in the steel wall should be discarded by the machine. It seems that the leaning machine cannot sufficiently distinguish between the two pipe types. It is envisaged to train two separate models, one for seamless pipe one for welded pipe, to circumvent this problem. Keeping in mind, that the total number of features was 15777, the total number of type one errors and their relevance to not seem too dramatic.

The influence of boxing: It was known from past experience that the boxing process had a considerable influence on the later result of data analysis. In MFL inspection it could even affect the calculated depth of the metal loss. For ultrasonic inspection data it should at least be examined, how the boxing can influence the classification result. Naturally this again mainly depends on the training sample. The boxing should in general always be done with similar parameters as the ones initially used to generate the training samples. Parameters influence things like connectivity between boxes, number and size. The left side of Figure 5 shows a metal loss type feature with an increasing box size. There are altogether 10 boxes with the current box always encompassing the previous one. The right side of Figure 5 shows the activation of the two feature types irrelevant and metal loss.
The intermediate size would yield a high activation for metal loss, while it is smaller for sizes too small or too high. The classification metal loss would be correct here. It should thus be ensured that either the boxing parameters are not changed considerably over time from one inspection to the next or that training samples of boxes generated with all kinds of parameters are included.

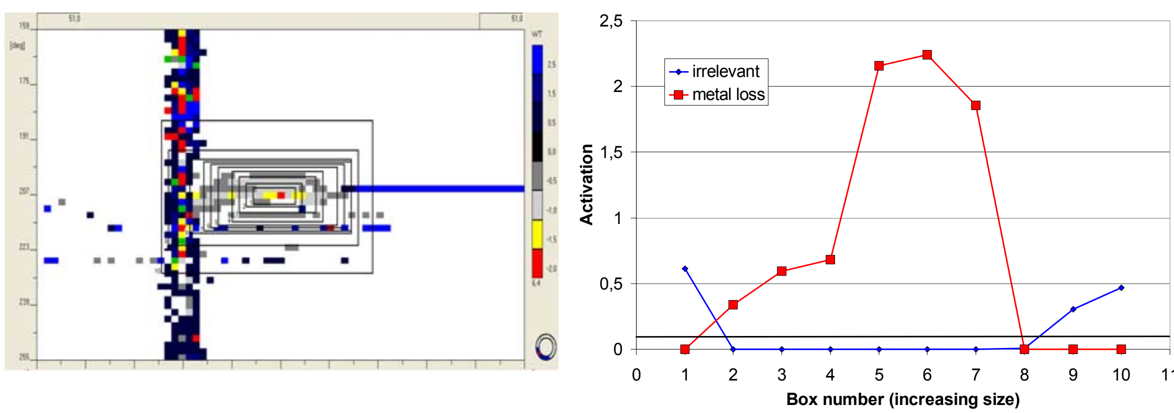

Figure 5: Effect of increasing box size on the activation of different classes. Due to the selection in the training samples there is an optimum size.

The handling of training data: The question on how the knowledge base is increased is a crucial one for the value of the system as a whole. Recent systems have been trained in the beginning and were then in use for many years. One aim of this implementation has been to make retraining as easy as possible. Of course the process of retraining has to be clearly defined to avoid any problems in keeping up with all earlier SVM models. The new training samples are extracted from the inspection work currently being processed. Once the manual verification of relevant and ambiguous features is done, the final classification is known. All features that have either been classified as ambiguous or that have been classified incorrectly are now candidates for the next enlarged training set.
This could be carried out every inspection. To retrain after every application of the automatic system would thus result in a very frequent process. The enlargement of the knowledge base would be very continuous. Also a frequent retraining necessitates the archiving of many models. It is always important that the deriving of the result can be traced back. Thus it is archived what inspection has been done with what model of the SVM. A less frequent renewal of the SVM-model is thus desirable. The database to store the data analysis results is thus set up to keep

the automatic as well as the manual results of classification. A flag can be set to mark a special feature as a training sample. In this manner the retraining can be done very quickly, and if a large improvement is anticipated, but it can also be done any time in the future, if the manpower is to be directed on other issues.

With increasing learning the number of training samples is growing. Every retraining requires the presence of the current set of training samples. By nature of the SVM only very few feature vectors will later be selected to be relevant (the support vectors). For instance the training set may include many similar shallow metal loss features. The training will determine the relevant features that describe the decision boundary. These features may only be a small number. A reduction of the training set, which keeps the knowledge base on the same level, can solve this problem. Methods to reduce the size of the training set are currently subject to active research [3].

Conclusions: Support vector machines represent a very efficient means of implementing artificial intelligence for the analysis of data obtained in pipeline inspection. The retraining and thus the maintenance of the knowledge base is considerably easier as compared to other methods. The possibility to retrain on demand, i.e. for special instances, where an instant improvement is anticipated, is now possible.

The successful application of the presented system will allow to speed up the analysis process for the generation of final and preliminary reports. The decision making will become reproducible and will be less dependent on human errors.

References:[1] R. Suna, K. Berns, K. Germerdonk, A.O. Barbian, *Pipeline diagnosis using backpropagation networks*, Neuro-Nimes, 1993, [2] Cristianini, Nello and John Shawe-Taylor (2003). *Support Vector and Kernel Methods*. In: Berthold, Michael and David J. Hands (Eds.): *Intelligent Data Analysis*. Springer-Verlag Berlin Heidelberg New York.],[3] M. Ziegenmeyer, *Optimierung und Anpassung der Support-Vektor-Klassifikation motiviert durch reale Diagnoaseanwendungen*, MasterThesis, FZI, Karlsruhe 2003