



TITLE STORY // MACHINE LEARNING

# ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING APPLIED IN COMPUTER AIDED ENGINEERING

Self-learning algorithms prove high potential in metamodeling for CAE-based applications. Especially if large amounts of data from simulation or experiments are available, they outperform “classic” regression methods.

## Introduction

Artificial Intelligence (AI) is the new shooting star in science and every industry has high hopes in using the “new” technology for almost everything. In this article, we would like to discuss what is “new” in AI for Computer Added Engineering (CAE) and what we can expect using AI-based technology for CAE applications. Starting with Wikipedia ([en.wikipedia.org/wiki/Artificial\\_intelligence](http://en.wikipedia.org/wiki/Artificial_intelligence)): “In computer science, artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals. Colloquially, the term “artificial intelligence” is used to describe machines that mimic “cognitive” functions that humans associate with other human minds, such as “learning” and “problem solving”. Continuing in related articles, we can extract that, like AI, different mathematical tools, which mimic natural intelligence by machine intelligence using computers, are understood. By searching for CAE applications, we will find several methods which are well known for a long time. This is mainly the methodology for solving optimization problems inspired by nature, such as evolutionary algorithms including genetic programming, swarm intelligence, as well as simulated annealing or, inspired by the human brain, so-called Artificial Neural Networks (ANN) are used.

Especially so-called Machine Learning (ML) algorithms extending ANN to deep forward neural or deep recurrent network as well as using support vector machines, which have made significant progress in the last decades, is indeed “new” in the context of AI for CAE. Taking advantage of that new technology and the availability of large data set and rising compute power, we expect that ML has a large potential for surrogate modeling. In purely mathematically driven surrogate models, often called meta-models, the learning process tries to understand how the response variability of



Fig. 1: Typical steps of machine learning algorithms (Source: [www.machine-learning-blog.com](http://www.machine-learning-blog.com))

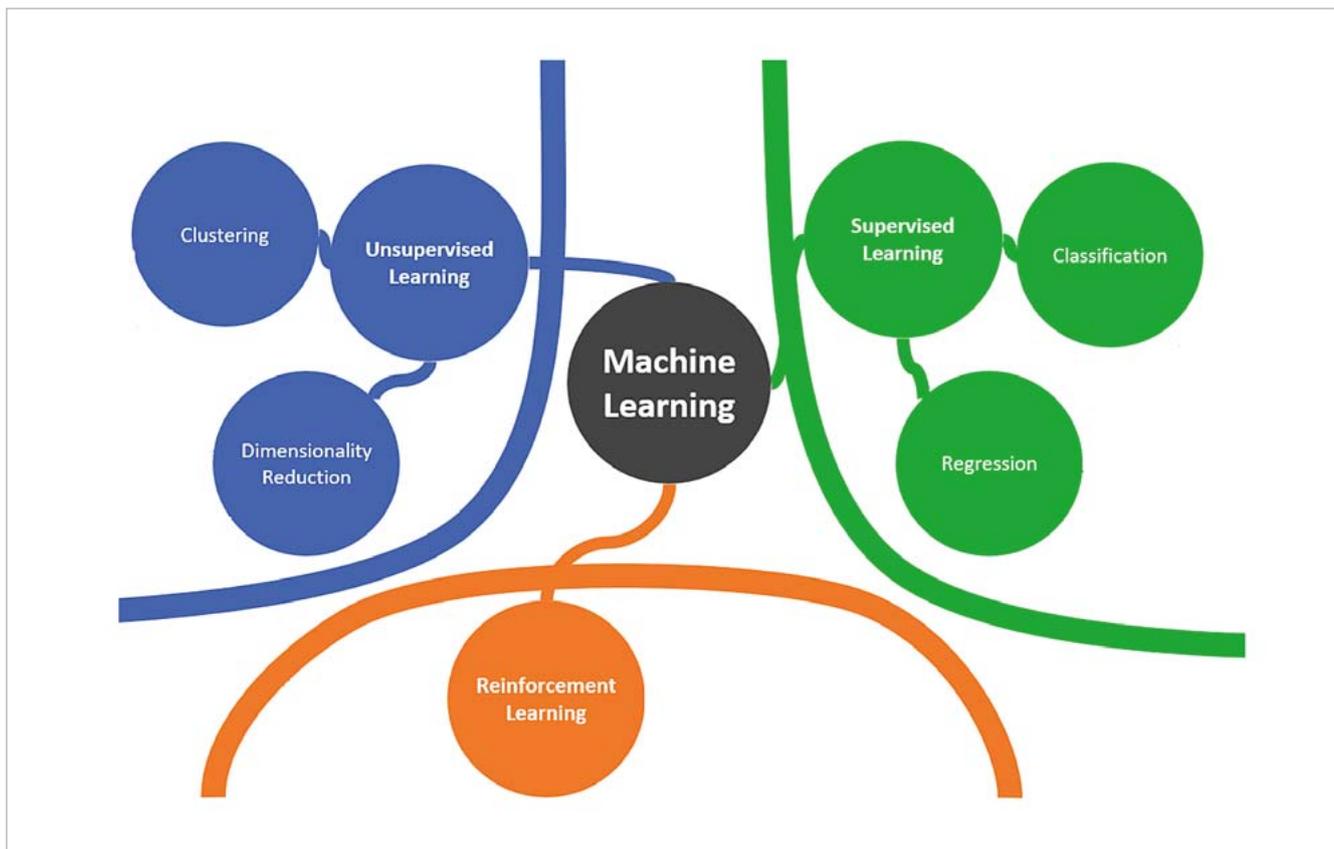


Fig. 2: Overview on different Machine Learning strategies (Source: [www.morethandigital.info](http://www.morethandigital.info))

a data set is correlated to variability of the inputs directly on the available data sets without further assumptions on the underlying physics. The required data may be collected from real measurements as applied for digital twins or in autonomous drive applications or may be obtained from simulations typically used in CAE based design.

Of course, using correlation and regression analysis for meta-modeling, the process of clean, train, test data and try to learn what is the best combination of dimensionality and basis function is also not really new. Therefore, this article will start with introducing classical regression methods for meta-modeling, discussing the problem of overfitting, which does not disappear with machine learning, define machine learning algorithms and finally present the integration of machine learning technology into Dynardo's software tools.

### Classical regression methods

In classical regression methods, a defined set of basis functions is used to set up a mathematical response surface model. Linear regression (Montgomery 2003) with linear or higher basis functions are a common basic approach, where a global set of basis functions is defined for a specific problem and for each basis function term the corresponding coefficient has to be calculated. Often this is done by a minimized least squares approach, in which the squared errors between data and model approximation are mini-

mized. If the chosen global basis terms are not suitable to represent a physical phenomenon, this obtained response surface function is even in cases with sufficient high number of data points often not acceptable accurate.

In contrast to the global approach, local response surface methods like Moving Least Squares, Radial Basis Functions, or Kriging use basis terms with local support functions. With the help of a scaling factor the area of this local support can be defined. All available data points are usually considered as support of these local basis functions. This means, with increasing the data amount, the approximation quality of these local models raises. But in case of noisy data points, caused by solver noise in CAE or measurement errors in experimental analyses, such local models may tend to overfitting, by representing artificial oscillations in the approximation function. The overfitting needs special attention in the training and in the quantification of the accuracy of the surrogate model.

In the classical response surface approach, usually a specific model type and basis terms are chosen and trained with the data and later the approximation quality is quantified by some error measure. Often the Coefficient of Determination (CoD) as a measure of the goodness of fit is used for this purpose. Unfortunately, measuring the accuracy of fit is only suitable in cases, when the underlying regression model is not arbitrary flexible to enable local overfitting. In case of a global linear regression, the CoD is suitable, if the number of unknown regression terms is much smaller

as the number of available data points. If local regression models are used, as MLS or Kriging the CoD measure usually becomes too optimistic. In cases with significant solver noise, the approximation function may be strongly distorted by local oscillations which imply artificial non-physical correlations. Therefore, measuring the goodness of fit is incapable to avoid the overfitting and hence measuring the forecast quality by means of independent data point sets becomes crucial.

Today, many more meta-model approaches are available and it is often not clear which one is most suitable for a given problem (Roos 2007). Another challenge of meta-modeling is the so-called “curse of dimensionality”, meaning, that there is a dramatic decrease in the quality of approximation for all meta-model types as the number of input variables increases. As a result, large number of samples are required to represent high-dimensional problems with sufficient accuracy, having at the same time the tendency to overfit the data as pointed out above. In order to overcome these problems, Dynardo developed the Metamodel of Optimal Prognosis framework (Most and Will 2008, 2011), which will be discussed in a later section. Prior to that, we would like to introduce different strategies of machine learning.

### Machine Learning

The machine learning approach can be seen as an extension of the classical response surface methods. Instead of choosing a predefined setup of basis function types and terms manually, this approach tries to learn the complexity of the required model type itself. The type of this automated learning now distinguishes between the different levels of machine learning.

In the **unsupervised learning** so-called unlabeled data, which contain only the information about the input values, are considered. In clustering approaches for example, this may help to find different regions in the variable space, where different data clouds are available, which might be treated differently. On the other hand, dimension reduction techniques are very promising, in cases where a large number of input variables, which show significant local or global dependencies to each other, needs to be considered. Exemplarily, the Principal Component Analysis method or related series expansion approaches are mentioned, which are very efficient to reduce a large input dimension to a few representative variables. Reduction strategies used in Statistics on Structures (SoS) can be labeled as unsupervised machine learning.

In the **supervised learning approach**, the labels of the data, often understood as response variables, which describe a phenomenon depending on the input variables, are represented by a set of mathematical surrogate functions. During the learning or training process, the complexity of the function set is automatically adjusted to the requirements represented by the data. For example, the Support Vector Classification and Regression, which uses local basis functions similar to Kriging, reduces the set of relevant ba-

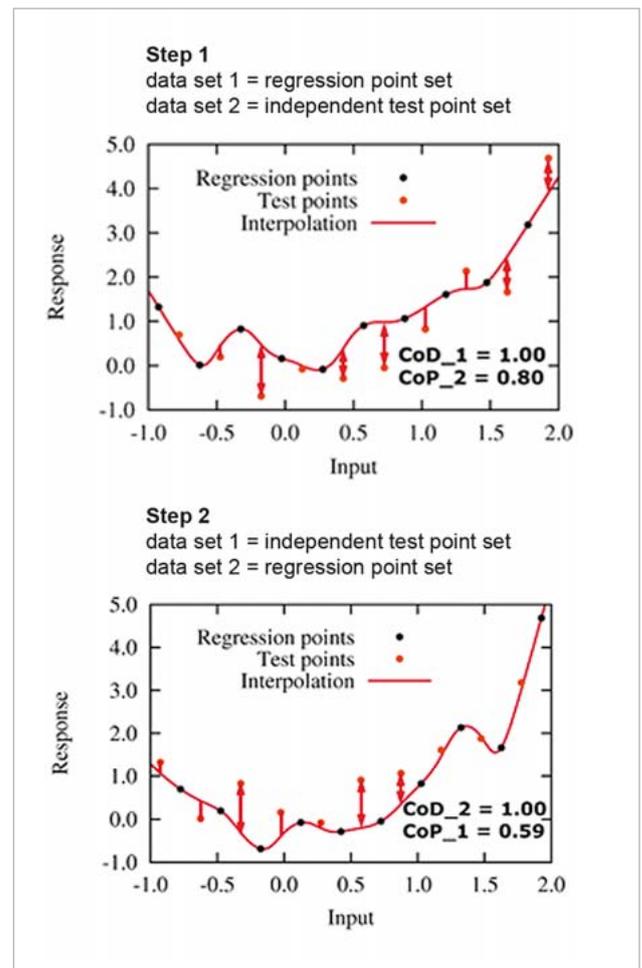


Fig. 3: The Coefficient of Prognosis as measure of the prediction quality based on independent test data points

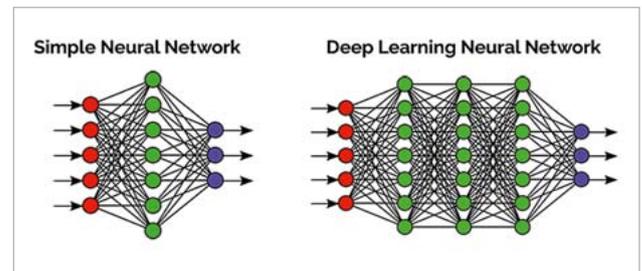


Fig. 4: Simple feed-forward neural network consisting of one hidden layer and deep learning network with several hidden layers

sis terms automatically to these support points, which are required to build the class separation or the response function sufficiently. This case is an example for automatic selection of necessary data points. In the chapter “Metamodel of Optimal Prognosis” Dynardo’s MOP technology is introduced and discussed as supervised machine learning approach with the focus on detecting the most important input variables as well as suitable regression functions.

Reinforcement learning approaches differ from supervised learning by the presence of labeled input/response pairs, and by sub-optimal pairs not being explicitly corrected. Instead the focus is on finding a balance between

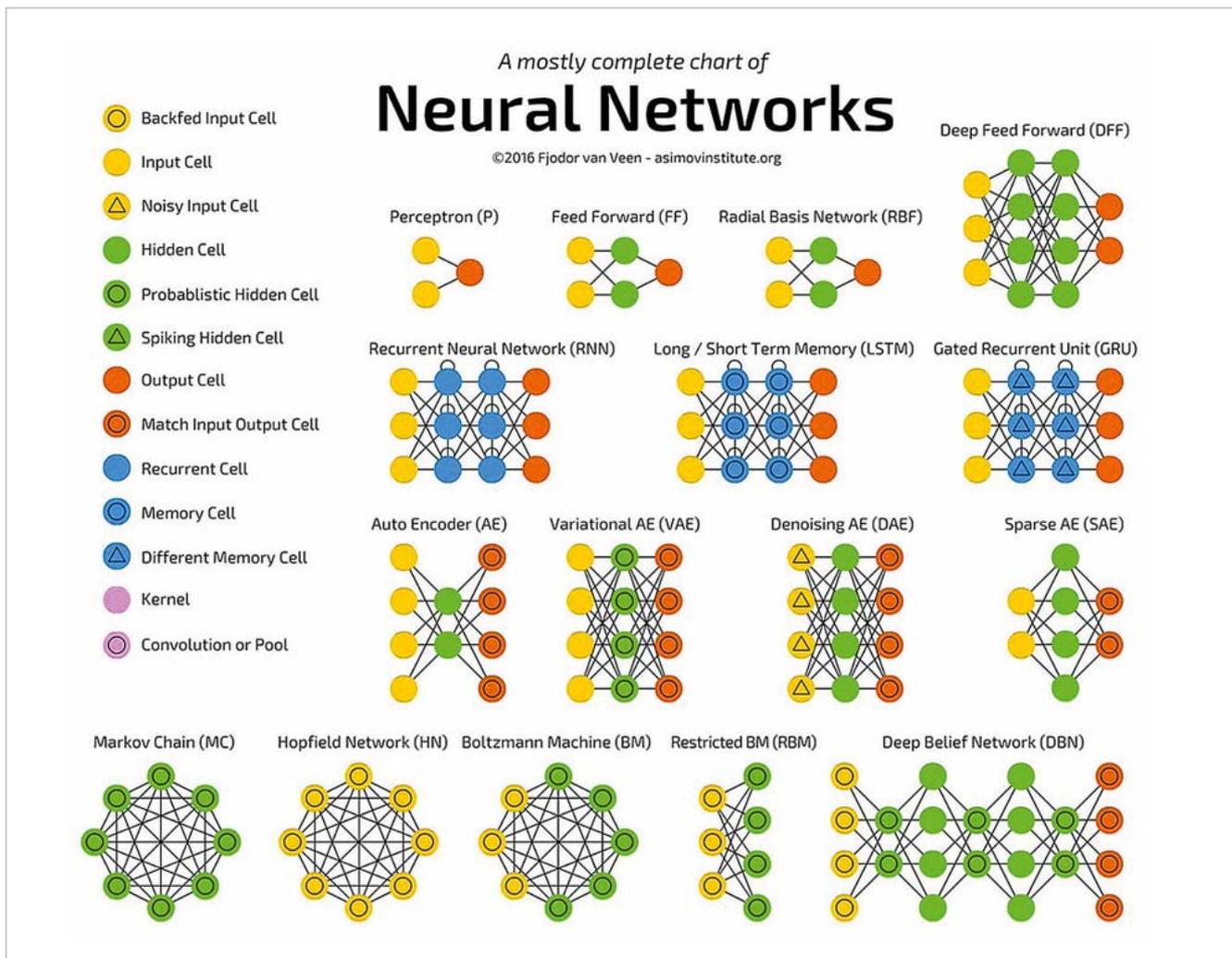


Fig. 5: Typical neural network types and architectures (Source: Fjodor van Veen, Asimov Institute, 2016)

exploration (of uncharted input territory) and exploitation (of current meta-models about input response correlation). Therefore, adaptive strategies are used to add additional support points in uncharted areas, respectively areas of interest to improve the forecast quality of meta-models. Dymardo's Adaptive Metamodels of Optimal Prognosis (AMOP) can be classified as a reinforced learning strategy.

### Metamodel of Optimal Prognosis

The Metamodel of Optimal Prognosis (MOP, Most & Will 2008, 2011) can be seen as an advanced machine learning approach with automatic feature detection. Based on the Coefficient of Prognosis (CoP), which measures the forecast quality of an approximation model instead of the goodness of fit, the MOP approach evaluates different regression methods with different basis terms. Based on advanced filter strategies, different input variable combinations, which span different subspaces, are analyzed for different available regression model types. As a result, the optimal input variable set including the optimal approximation technique is obtained, which reaches the largest CoP value for a given support point data set.

One key point in this approach is the CoP, which allows to quantify the prediction quality of the approximation model independently of the model type. This measure is applicable also for interpolation models with perfect goodness of fit. The CoP estimates the fraction of explained variation in the prediction of the model. The residuals are calculated by using an independent test point set to estimate the approximation errors at test points. Based on the cross-validation principle, the prediction residuals are obtained for each data point set by a systematic exchange of regression and test points. With help of this procedure, an automatic selection of regression model types and variable subspaces is possible within the machine learning process.

Figure 3 shows a data set, which is splitted 50/50 to test and regression points in case 1 and vice versa in case 2. Unless advanced regression models are able to show a perfect fitting quality (CoD), the forecast quality (CoP) quantifies the true ability of the models to represent the data set response value variation. The difference between CoD and CoP quantifies the overfitting. It should be noted that in the test case 23% of artificial noise is represented. By using more data points, the CoP measure of any representative split will converge to 77%.

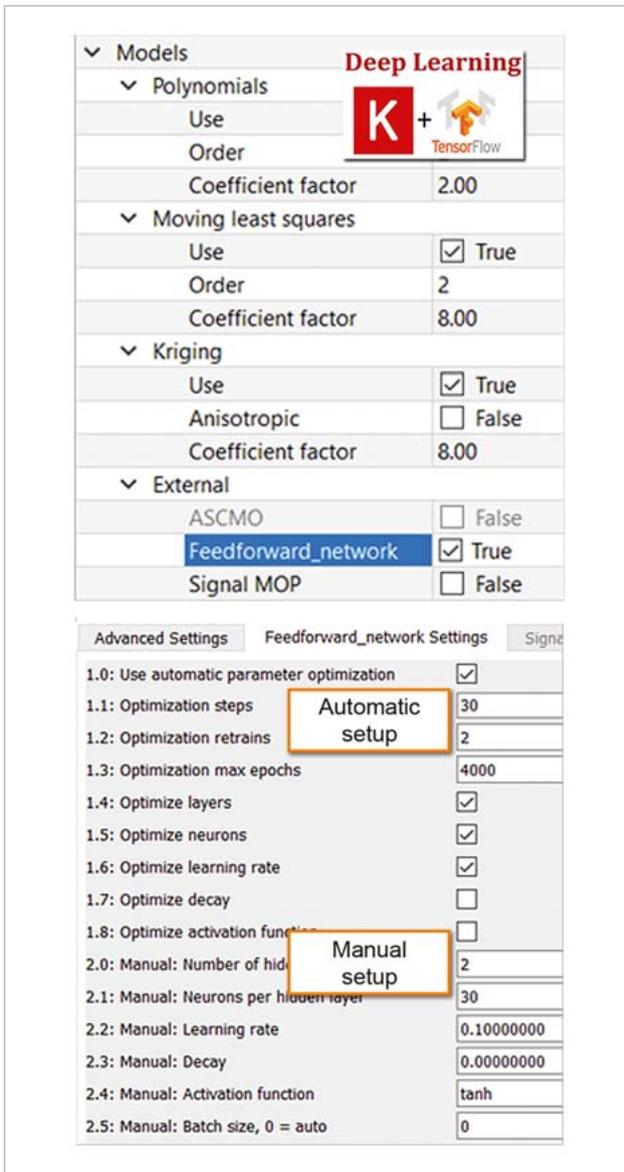


Fig. 6: Set up of optiSlang Deep Learning Extension models in the MOP settings (top: selection of feed-forward model for the MOP competition, bottom: automatic and manual network setup)

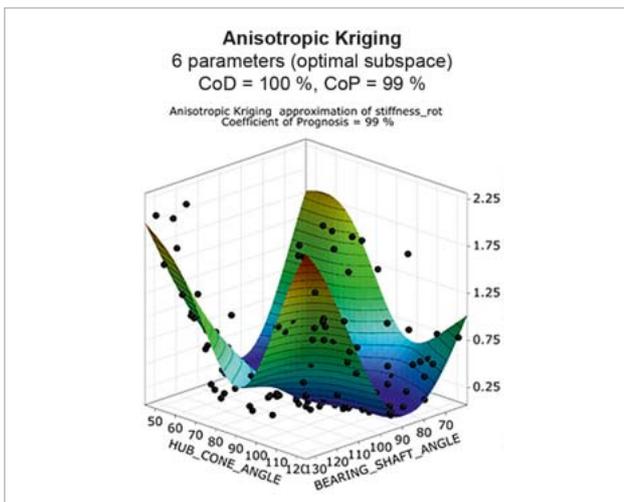


Fig. 7: Rotational stiffness of a turbine impeller approximated by Kriging

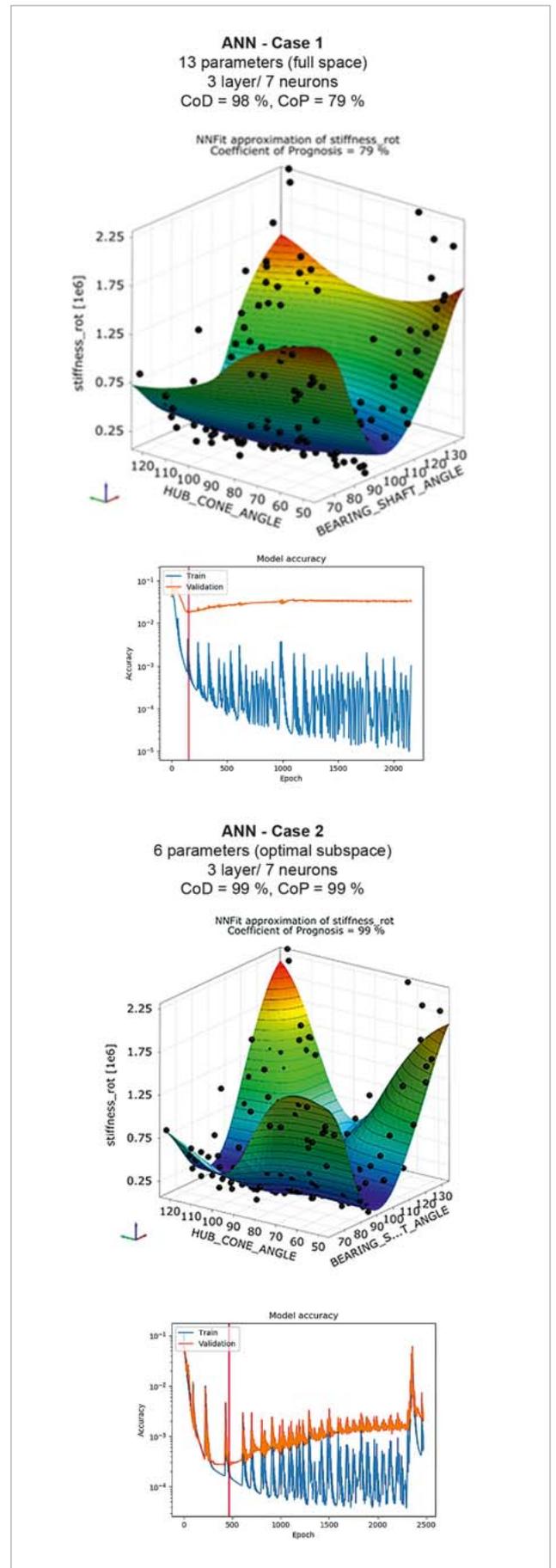


Fig. 8: Rotational stiffness of a turbine impeller approximated by a feed-forward network within the full parameter and the reduced optimal parameter space

Samples	Kriging		Deep Learning Extension		
	Training	CoP	Hidden neurons	Training	CoP
100	2 sec	99.99 %	3 x 10	3 min	99.73 %
500	3 min	99.99 %	3 x 20	7 min	99.96 %
2000	330 min	99.99 %	4 x 30	12 min	99.98 %
10000	not possible		4 x 50	72 min	99.99 %

Table 1: Comparison of numerical performance between Kriging and neural networks on a five-dimensional analytical benchmark function using computation on CPU only

### Artificial Neural Networks and Deep Learning

Artificial neural networks are a special type of machine learning methods, which mimic the learning process of the human brain: based on simple neuron activation functions, sophisticated phenomena can be learned just by increasing the complexity of the network with more and more neurons and connections. So-called deep learning neural networks consist usually of several layers of neurons with non-linear activation functions, whereby each neuron of each layer is itself connected to all neurons of the previous and the following layer. With increasing number of neurons, the complexity of the overall surrogate function increases significantly. Since the training algorithms are mainly based on simply update rules using stochastic optimizers, even such sophisticated deep learning networks can be efficiently trained. Especially, during the last ten years, the efficiency of the training methods could be improved dramatically. This opened the door for more and more complex network types and architectures and therefore for more applications.

In comparison to classical regressions methods, whose number of data points usually increases the training effort exponentially, the training of deep learning networks is still efficient for large data sets. Due to the simplicity of the training, parallelization techniques can be applied easily.

In cases of a sufficient amount of data, a deep learning network structure is able to detect important features itself automatically during the training procedure. In opposite cases, where the amount of data is limited, the reduction of input variables to only the important subset in the MOP workflow can significantly improve the approximation quality of a neural network similarly to classical regression methods. Usually, in such cases, the different approximation models show a similar performance. In our experience, the deep learning networks are a very powerful extension to the classical regression methods, especially if large amount of data are available.

### optiSLang's Deep Learning Extension

With the optiSLang deep learning extension, deep learning neural networks can be considered in the model testing process of the Metamodel of Optimal Prognosis. Based on the CoP, the available models are tested and compared. Finally, the best model type is selected. Especially, for large

data sets with more than 1000 samples, a significant speed up for the training is possible compared to classical methods as Kriging. Actually, the deep learning extension contains feed-forward deep learning models with automatic configuration of the network topology. Within the extension, the analyzed different model architectures are initialized and trained using the Tensorflow library, which can be efficiently parallelized on CPU and GPU environments.

### Example

The rotational stiffness of a turbine impeller was analyzed. Therefore, 176 samples of a Latin Hypercube Sampling with 13 geometry parameters as inputs have been considered. The response values, as rotational and axial stiffnesses as well as performance measures like isotropic efficiency and mass flow and even life time estimates have been obtained for each parameter combination by a time-consuming multi-disciplinary analysis including computational mechanics, fluid dynamics and fatigue analysis. That industrial example was used to test the performance and accuracy of the optiSLang deep learning extension. In Figure 7 the reference solution using the MOP dimension reduction and Kriging approximation is shown. The best possible global forecast quality is 99%. In Figure 8 the approximation functions for the rotational stiffness are shown for a neural network approximation considering all 13 parameters on the first case and only the most important 6 geometry parameters in the second. In case one Figure 8 clearly indicates that the highly nonlinear behavior of the response cannot be represented sufficiently. Due to the flexibility of the model, the accuracy of fit, measured by the CoD, is almost perfect, while the forecast quality, quantified by the CoP, is significantly less. Consequently, the deep learning training, which uses early stopping based on a validation data set to avoid overfitting, results in a rather weak trained model in the full space. This phenomenon is critical especially for smaller data sets. In such cases the automatic variable filtering of the MOP approach is very promising, as shown in this example. The forecast quality could be increased from 81% to 98% by training the neural network with the same topology but in the reduced optimal subspace.

As already pointed out in chapter "Artificial Neural Networks and Deep Learning", in case of small data sets

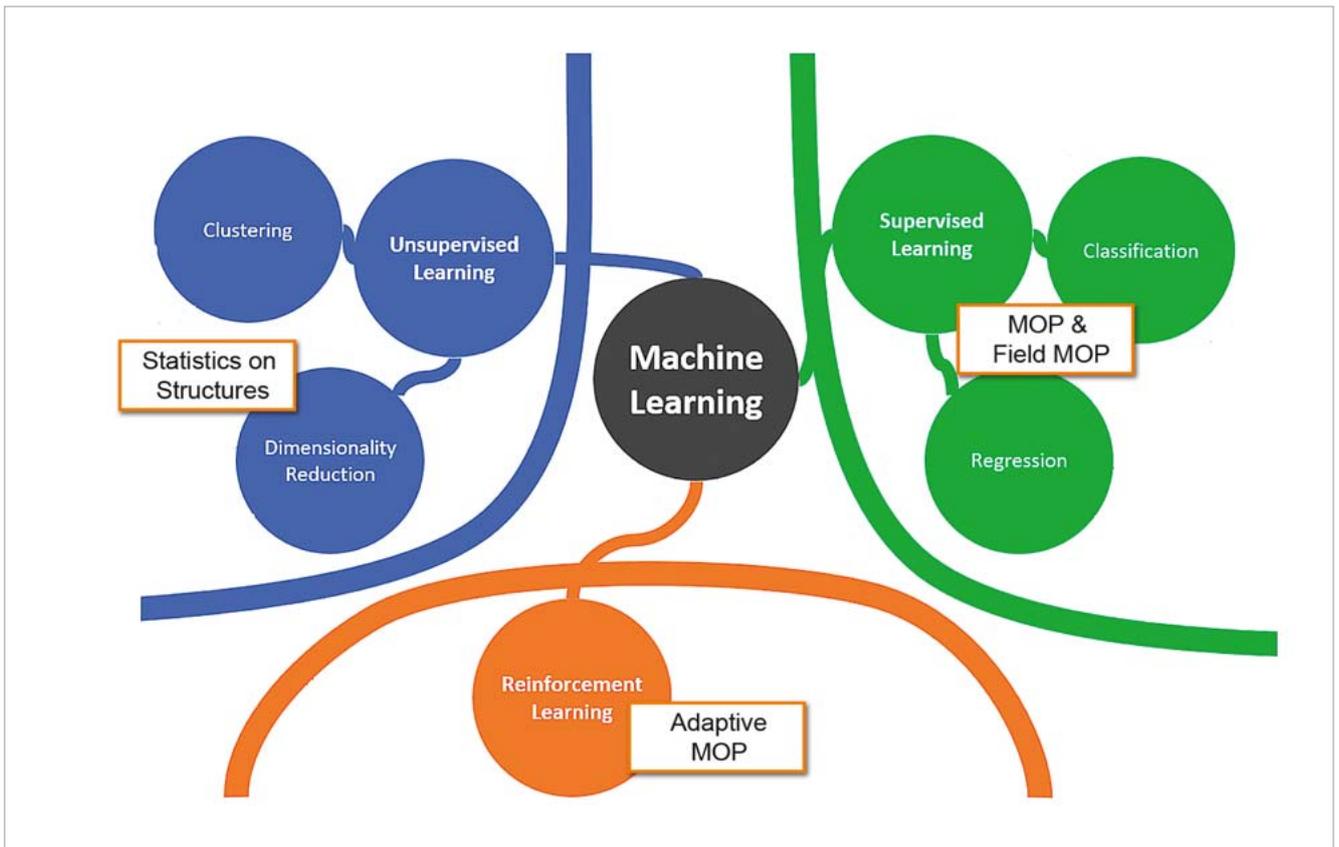


Fig. 9: Representation of different types of machine learning in dynardo technology (Source: www.morethandigital.info)

(50...200) the AI-based deep learning networks show a similar performance to MLS or Kriging in the MOP workflow. In our experience, the deep learning networks become a very powerful extension to the classical regression methods, especially if large amount of data are available. Table 1 (see previous page) shows the comparison of numerical performance using a five-dimensional test example of a non-linear function. It can be clearly seen, that Kriging, which is able to reproduce the nonlinearity, will fail because of its numerical effort when the available numbers of sample rise.

## Summary

This article discussed the relevance of “new” AI-technologies in the context of CAE applications. Fortunately, the CAE community in case of AI-based machine learning algorithms can take benefit of the enormous investments over the last decade in implementing public available libraries like TensorFlow from Google or Microsoft Cognitive Toolkit into CAE software applications. The article discusses in detail the MOP framework as supervised machine learning approach and the integration of feedforward neural networks into the MOP competition for the best possible meta-model. Within the MOP workflow, overfitting by using ANN is avoided due to rigorous testing of the forecast quality with help of the CoP. Users can investigate and quantify the global and local forecast quality of an ANN directly in comparison to classic meta-modeling techniques. The in-

terfacing of external machine learning algorithms to the MOP workflow is open to any third-party machine learning algorithm.

Right now, we are working on extensions to non-scalar meta-models, which will open the competition to Dynardo internal solutions for field MOP using Statistics of structures (SoS) as well as to adaptive scalar and field meta-modeling techniques to improve meta-modeling quality by optimally placed additional simulated data points (Adaptive Metamodel of Optimal Prognosis, AMOP).

## Authors //

Thomas Most, Johannes Will, Jonas Rotermund, Lars Gräning (Dynardo GmbH)

## References

- D. C. Montgomery and G. C. Runger (2003). Applied Statistics and Probability for Engineers. John Wiley & Sons, third edition
- Most, T. and JWill (2008). Metamodel of Optimal Prognosis - an automatic approach for variable reduction and optimal metamodel selection, Proceedings Weimarer Optimization and Stochastic Days 5.0, Weimar, Germany
- Most, T. and J. Will (2011). Sensitivity analysis using the Metamodel of Optimal Prognosis, Proceedings Weimarer Optimization and Stochastic Days 8.0, Weimar, Germany
- Roos, D., T. Most, J. F. Unger, and J. Will (2007). Advanced surrogate models within the robustness evaluation. Proceedings Weimarer Optimization and Stochastic Days 4.0, Weimar, Germany
- Russell & Norvig (2009). Artificial Intelligence: A Modern Approach. Prentice Hall