COMPARISON OF MACHINE LEARNING ALGORITHMS RANDOM FOREST, ARTIFICIAL NEURAL NETWORK AND SUPPORT VECTOR MACHINE TO MAXIMUM LIKELIHOOD FOR SUPERVISED CROP TYPE CLASSIFICATION

I. Nitze [a], U. Schulthess [b], H. Asche [c]

a,c Universität Potsdam, Institut für Geographie, 14476 Potsdam – ingmarnitze@gmail.com, gislab@uni-potsdam.de
b 4DMaps, 10405 Berlin - uschulthess@4dmaps.de

KEY WORDS: Crop Classification, Machine Learning Algorithms, Support Vector Machine, RapidEye

ABSTRACT:
The classification and recognition of agricultural crop types is an important application of remote sensing. New machine learning algorithms have emerged in the last years, but so far, few studies only have compared their performance and usability. Therefore, we compared three different state-of-the-art machine learning classifiers, namely Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest (RF) as well as the traditional classification method Maximum Likelihood (ML) among each other. For this purpose we classified a dataset of more than 500 crop fields located in the Canadian Prairies with a stratified randomized sampling approach. Up to four multi-spectral RapidEye images from the 2009 growing season were used. We compared the mean overall classification accuracies as well as standard deviations. Furthermore, the classification accuracy of single crops was analysed. Support Vector Machine classifiers using radial basis function or polynomial kernels exhibited superior results to ANN and RF in terms of overall accuracy and robustness, while ML exhibited inferior accuracies and higher variability. Grassland exhibited the best results for early-season mono-temporal analysis. With a multi-temporal approach, the highest accuracies were achieved for Rapeseed and Field Peas. Other crops, such as Wheat, Flax and Lentils were also successfully classified. The user’s and producer’s accuracies were higher than 85 %.

1. INTRODUCTION

Crop type classification is an important application of remote sensing. It is potentially much faster, more accurate and therefore more cost effective than conventional methods of generating regional crop area estimates. Crop type information at the field level can be used for agricultural surveys, subsidy control or, as auxiliary information for the prediction of crop yield and shortages thereof.

In the following paper we are comparing the machine learning classifiers Random Forest (RF) (Breiman, 2001), Artificial-Neural-Network (ANN) (Rosenblatt, 1958; Rumelhart et al., 1986), and Support-Vector-Machine (SVM) (Cortes & Vapnik, 1995). As a reference, we are also including the Maximum Likelihood (ML) algorithm, the most popular traditional supervised classification method. So far, the machine learning algorithms have not been widely used for crop classification, and as to our knowledge, their performance in this type of application has not been thoroughly compared.

2. STATE-OF-THE-ART

Since the dawn of remote sensing, numerous studies on crop-type classification have been published. Either optical or the combination of Radar and optical data were used as primary data sources.

In recent studies (Yang et al., 2011; Dixon & Candade, 2008) the superiority of non-parametric machine-learning algorithms to parametric classifiers, such as nearest neighbour or ML, has been described. Classification accuracies of Decision Trees such as RF, Artificial Neural Networks and Support Vector Machine were found to be similar. Artificial-Neural-Network and SVM achieved similar results in a land-cover classification study on Landsat TM data carried out by Dixon & Candade (2008), whereas ML performed significantly worse. In a comparison of Decision Tree (DT), ANN and ML, Pal & Mather (2003) reported non-significant differences in classification accuracy between the former two, whereas the manual work- and computational time effort turned out to be much more intensive for ANN. A land-cover classification with ANN, SVM, DT and ML published by Huang et al. (2002) resulted in higher accuracies of ANN and SVM as compared to DT. Nonetheless DT performed much faster with a calculation time of minutes compared to hours and days respectively for SVM and ANN. Hence, the major disadvantage of the machine learning classifiers is that their computational complexity is higher compared to traditional supervised methods, such as ML or Nearest Neighbor, but they also differ strongly among each other.

However, as to our knowledge so far no crop classification study compared the main machine-learning algorithms RF, ANN and SVM altogether.

3. DATA AND METHODS

3.1 Data

The classification was performed on a multi-temporal set of optical RapidEye images. They have a ground sampling distance of five meters and cover the optical electro-magnetic spectrum in five bands: blue, green, red, red-edge and near-infrared (RapidEye, 2011).

The study area of 20 by 25 km was located around the municipality of Indian Head, in south-eastern Saskatchewan, Canada. It contained 512 agricultural fields of known crop or cultivation types grown during the summer of 2009 (cf. Fig. 1).

The geo-referenced field boundaries, including crop type, seeding date and many more types of information, were originally collected for the ESA AGRISAR campaign 2009 (ESA, 2009).

After excluding very small classes and merging semantically similar classes, 10 different crop types were left for supervised classification. They consist of *Wheat, Barley, Rapeseed, Oats, Field Peas, Lentils, Canary Seed, Flax, Grassland* and *Fallow* (cf. table 1). Four cloud free images had been acquired on June 2, August 10, August 25 and September 5. The study area was covered by RapidEye tile IDs 1363418 and 1363419.

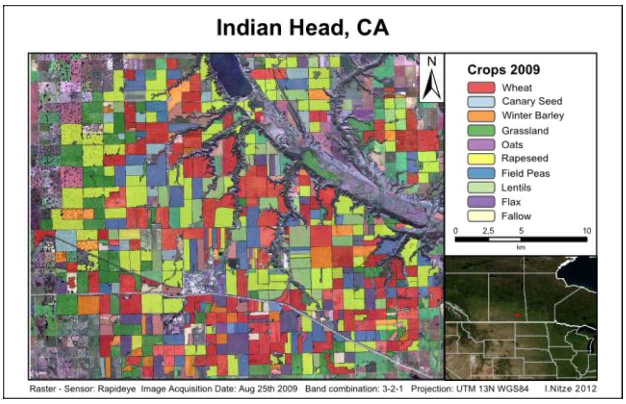

Figure 1: Overview of study area Indian Head including field boundaries with crop types

Crop type	# of fields
Wheat	161
Rapeseed	136
Grassland	79
Field Peas	52
Barley	40
Lentils	38
Flax	37
Oats	30
Fallow	19
Canary Seed	18

Table 1: Cultivated crops with number of fields in study area Indian Head (CA)

3.2 Methods

The workflow can be divided into two major steps: data pre-processing and classification.

3.2.1 Pre-processing: At first, the spatial co-registration of field-boundaries to the images was visually checked and if necessary, corrected. In order to reduce the influence of mixed pixels at the edge of the fields, the field-boundaries were buffered by ten meters to the inside.

The RapidEye data were processed in several steps. At first, the two image tiles of each acquisition date were mosaicked. Afterwards a basic atmospheric correction was performed on the mosaic. This includes radiance to top-of-atmosphere reflectance conversion as well as dark-object-subtraction (Chavez, 1996). Moreover, five different vegetation indices were calculated: ground cover (Maas & Rajan, 2008), NDVI (Rouse et al., 1973), MTVI2 (Haboudane et al., 2004), NDVIRE (Barnes et al., 2000), and MTCI (Dash & Curran, 2007). They were calculated as additional sources of information in order to increase overall classification accuracies.

In the next step, the spectral information as well as vegetation indices were extracted: the median of each of the five spectral bands and vegetation indices inside each buffered field boundary was calculated and saved to the attribute table of the vector containing the files field boundaries. This resulted in ten additional attributes per image date, 40 in total. We chose the median over mean due to a higher stability against outliers, which may occur due to anomalies and small water bodies within the crop fields.

3.2.2 Classification: Weka 3.6, a collection of machine learning algorithms for data mining tasks was used for the analysis (Hall et al., 2009). It is an Open-Source-Software-Package, which can be used for classification and includes a built-in validation function.

In order to assess the change in classification accuracy with time, four different image-date combinations were used. Starting with the first image, one image was added in each subsequent analysis, until all images in the study area were used in conjunction. This resulted in four datasets.

The following classifiers were used: Naïve Bayes for ML, Random Forest (RF), Multi-Layer Perceptron in case of ANN, and LibSVM for Support Vector Machine. For the latter radial-basis-function (RBF) and polynomial kernels (POLY) were applied to the classification, hence five different classifiers were used in total.

Parameter optimization for the machine learning classifiers was carried out with GridSearch for SVM-RBF, SVM-POLY and ANN while CVParameterSelection was used for RF. For SVM-RBF parameters *gamma* and *C* were optimised in exponential steps of 0.5 between 10^{-5} and 10 as well as 10^{-1} and 10^{5}.

The resulting values were then applied to the parameter search of SVM-POLY. Here the *polynomial degree* and *coef0* parameters were searched between 2 and 6 as well as 10^{0} and 10^{6} respectively.

The optimization of Artificial Neural Network includes two different steps: choice of architecture and parameter search. Initially we tested different configurations with one or two hidden layers and a variable number of hidden neurons. After finding the most successful and robust architecture, the number of neurons in the hidden layer equals number of attributes, the parameters *learning rate* and *momentum* were both optimized in steps of 0.1 between 0 and 1.

Finally the best *number of trees* between 100 and 1000 was determined using the *CVParameterSelection*-Function for Random Forest Classifier.

3.2.3 Experiment 1: Assessment of overall classification accuracy. This task was performed with the *WEKA Experimenter*. For each dataset the beforehand optimized classifiers (cf. Table 2) were run 100 times each with an automatic random stratified selection of training and test sets, but same conditions for all classifiers. The splitting ratio was set to 80 % test and 20 % training data. All iterations were automatically validated by the software and exported to a csv-file containing different quality measures. These include overall accuracy, kappa as well as training and testing time. However, single class information was not included. The overall accuracy was then analysed using statistical parameters such as mean and standard deviation.

	Date 1	Date 2	Date 3	Date 4
Parameter	ANN			
LearnRate	0.6	0.6	0.7	0.4
Momentum	0.3	0.1	0	0.2
Parameter	RF			
nTrees	500	400	100	600
Parameter	SVM			
Gamma	0.01	1	0.3	0.3
C	300000	30	30	30
Degree	3	4	2	2
Coef0	1	10	1	3

Table 2: Calculated optimized parameters per classifier and dataset.

With these measures a comparison of classification success, using the mean overall accuracy, and robustness, using standard deviations, was carried out. Finally the statistical significance of the different results was assessed by using a paired t-test at a significance level of 0.05.

3.2.4 Experiment Setup 2: Assessment of classification accuracy for single crop types as a function of type of classifier and number of images. For this purpose all datasets were classified five times with each classifier using 5-fold cross-validation, in correspondence to the split-ratio in Experiment 1. The median result of each setup, run in *WEKA Explorer*, was then used as the reference result. Producer's accuracy, user's accuracy and F-Measure (cf. Equation 1) were used for quality assessment, whereas the error matrix provides information about class confusion.

$$F = \frac{2 \bullet PA \bullet UA}{PA + UA} \quad (1)$$

PA: producer's accuracy; UA: user's accuracy

4. RESULTS

4.1 Overall classification accuracies

Overall classification accuracy varied considerably among the classifiers used. Furthermore, the number of images used for the classification greatly affected classification accuracy. As shown in Figure 2 and Table 3 SVM-RBF exhibited the highest mean overall accuracies. The largest margin between different methods could be observed in mono-temporal analyses, using the first image only. SVM-RBF and SVM-POLY produced nearly identical results with 68.6 and 68.4 % respectively. Notably lower accuracies were observed using ANN and RF. The former achieved 61.8 % and the latter 55.8 % overall accuracy. According to the t-test the differences are statistically significant between SVM, ANN, and RF. With only 45 % overall accuracy, ML exhibited by far the worst results, which were also statistically different from the other ones.

# of images	1	2	3	4
Classifier	Mean overall accuracy [%]			
ANN	61.8	80.2	87.4	87.1
RF	55.8	82.3	86.6	87.4
SVM-RBF	68.6	82.2	88.0	88.1
SVM-POLY	68.4	80.3	87.7	87.8
ML	45.0	75.8	79.1	78.9

Table 3: Mean overall accuracies of all classifiers depending on number of image acquisitions in study area Indian Head.

With the addition of a second coverage on August 5 the mean overall accuracies were raised by 12 to 31 %. The differences in classification performance among the used classifiers became much narrower. As exhibited in Table 3, RF and SVM-RBF accomplished nearly identical results with 82.2 and 82.3 % mean overall accuracy. However their stability, as assessed by their standard deviations (STD) varied slightly (cf. Table 4). Random Forest showed slightly more stable results with a STD of 2.38 % versus 2.65 % for SVM-RBF. The classification accuracies of ANN and SVM-POLY exhibited rather similar characteristics, but displayed statistically significant lower accuracies by 2 %, as well as slightly higher STDs than SVM-RBF and RF. Maximum Likelihood showed again inferior results with a mean overall accuracy of 75.8 % and a STD of 3.11 %.

With the addition of a third image from August 25 the classification results were further improved by 3 to 8 %. The machine learning algorithms still outperformed ML by 7.5 to 9 % (cf. Table 3). SVM-RBF, again, reached the best mean accuracy with 88 % and the highest stability. Close second and third, SVM-POLY and ANN performed slightly worse at 87.7 and 87.4 % respectively, whereas RF achieved 86.6 %.

Adding the fourth and last satellite image from early September did not necessarily improve mean overall classification accuracies (cf. Table 3). Furthermore, in 2 cases a decrease in classification performance of 0.2 and 0.3 % for ML and ANN was observed. Both SVM classifiers exhibited near constant values, while the performance of RF was increased by 0.8 %. The stability of the classification results improved considerably to STDs of just over 2 %, with the exception of ML, which had a STD of 2.82 %.

# of images	1	2	3	4
Classifier	Standard deviation [%]			
ANN	3.89	3.17	2.69	2.04
RF	3.11	2.38	2.40	2.10
SVM-RBF	2.82	2.65	2.39	2.05
SVM-POLY	2.67	2.82	2.49	2.17
ML	4.03	3.11	2.81	2.82

Table 4: Standard deviations of all classifiers depending on number of image acquisitions in study area Indian Head (CA).

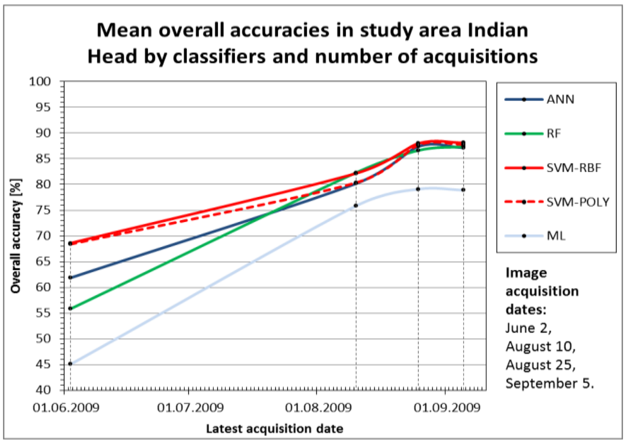

Figure 2: Mean overall classification accuracies achieved at Indian Head (CA) by five different algorithms as a function of number of images used.

4.2 Calculation complexity:

Execution times varied greatly among different classifiers and number of acquisitions (cf. Table 5). Artificial Neural Network took the longest time for training, with an average of 7.7 to 15.1 seconds training time per classification. With only one acquisition the computation times of the remaining machine learning algorithms resembled those of ANN.

With more acquisition dates and thus more complex datasets the calculation times were shorter and more diverse among RF, SVM-RBF and SVM-POLY. With 1.1 to 6.2 seconds RF is much more computationally expensive than both SVM classifiers, which showed similar durations. Their average computation time of the training stage oscillated around 0.3 seconds with the exception of SVM-POLY.

At two images with 0.684 seconds. ML featured by far the lowest computational cost with marginal time expenditures of 0.004 to 0.007 seconds per training stage.

# of images	1	2	3	4
Classifier	Training time per classification [sec]			
ANN	7.657	18.253	20.486	15.145
RF	8.765	4.002	1.134	6.205
SVM-RBF	9.631	0.281	0.335	0.292
SVM-POLY	8.452	0.684	0.313	0.296
ML	0.004	0.003	0.007	0.005

Table 5: Mean computation time used for the training of five different classification algorithms at Indian Head (CA).

In comparison to average training times, the testing or classification stage was generally performed much faster. For ANN the differences are particularly distinctive, while ML required slightly more time for testing than training. However both classifiers required only fractions of a second for the application of the learned models (cf. Table 6). Both SVM classifiers exhibited slightly longer testing times, but still no more than 0.039 seconds. With average testing times between 0.011 and 0.175 seconds RF usually needed more time than the remaining classifiers. The duration seemed to be strongly correlated to the number of used trees, which is paralleled to training times, where the usage of three images required the least and the usage of one classifier the most computational effort.

# of images	1	2	3	4
Classifier	Testing time per classification [sec]			
ANN	0.002	0.009	0.008	0.003
RF	0.175	0.065	0.011	0.083
SVM-RBF	0.035	0.018	0.030	0.039
SVM-POLY	0.018	0.014	0.019	0.020
ML	0.051	0.007	0.015	0.017

Table 6: Mean computation time used by five different algorithms for the crop classification at Indian Head (CA).

4.3 Single class results:

The classification results for individual crop types exhibited different kinds of behaviour in terms of classification accuracies, depending on number and dates of acquisitions as well as utilized classifier.

Using only one satellite image from June 2nd only *Grassland* could be safely classified with accuracies of around 90% (cf. Table 7). All classifiers besides RF featured superior results of producer's accuracies as compared to user's accuracies in excess of up to 9.7 %. Other notable results were obtained for the classes *Rapeseed* and *Wheat.* The former reached F-Measures of up to 87.4 % with SVM-RBF, but showed large differences between producer's and user's accuracies. Wheat showed similar trends, but generally lower accuracies. Classifier dependant results could be further observed for *Fallow* where ANN outperformed the other classifiers, and for *Field Peas* for which both SVM classifiers achieved the highest accuracies, with respectable F-Measures of more than 70 %. The error rates for the remaining classes were high.

The most notable class confusions were observed among the different cereal types *Wheat*, *Barley* and *Oats*. Other crops were also misclassified as *Wheat* or *Rapeseed*. In summary, large classes were usually over-classified while small classes were under-represented during classification with machine-learning techniques. This behaviour is backed by the already mentioned differences in producer's and user's accuracy.

Classifier	ANN			RF			SVM-RBF		
Measure [%]	PA	UA	F	PA	UA	F	PA	UA	F
Crop Type									
Wheat	56.3	80.1	66.2	55.5	72.0	62.7	58.4	92.5	71.6
Barley	28.6	10.0	14.8	21.7	12.5	15.9	0.0	0.0	0.0
Oats	33.3	6.7	11.1	28.6	6.7	10.8	20.0	6.7	10.0
Rapeseed	64.7	90.4	75.5	60.6	73.5	66.4	81.5	94.1	87.4
Canary Seed	25.0	5.6	9.1	28.6	11.1	16.0	0.0	0.0	0.0
Field Peas	59.5	42.3	49.4	51.1	44.2	47.4	72.0	69.2	70.6
Lentils	36.8	18.4	24.6	22.9	21.1	21.9	53.8	36.8	43.8
Flax	28.6	16.2	20.7	26.1	16.2	20.0	45.5	27.0	33.9
Grassland	94.7	89.9	92.2	88.9	91.1	90.0	95.8	86.1	90.7
Fallow	80.0	63.2	70.6	53.3	42.1	47.1	67.4	57.9	61.1
Classifier	SVM-POLY			ML					
Measure [%]	PA	UA	F	PA	UA	F			
Crop Type									
Wheat	59.3	95.0	73.0	55.3	45.3	49.8			
Barley	0.0	0.0	0.0	0.0	0.0	0.0			
Oats	33.3	6.7	11.1	25.0	16.7	20.0			
Rapeseed	80.4	87.5	83.8	47.1	48.5	47.8			
Canary Seed	0.0	0.0	0.0	6.1	16.7	9.0			
Field Peas	78.3	69.2	73.5	46.4	50.0	48.1			
Lentils	46.4	34.2	39.4	31.3	52.6	39.2			
Flax	44.8	35.1	39.4	12.8	16.2	14.3			
Grassland	93.2	87.3	90.2	93.2	86.1	89.5			
Fallow	55.0	57.9	56.4	33.3	47.4	39.1			

Table 7: Median classification accuracies for single crop types depending on classification algorithm used for the study area at Indian Head (CA). One satellite image from June 2 was used. PA: producer's accuracy; UA: user's accuracy; F: F-Measure.

Classifier	ANN			RF			SVM-RBF		
Measure [%]	PA	UA	F	PA	UA	F	PA	UA	F
Crop Type									
Wheat	84.4	90.7	87.4	78.9	95.7	87.0	85.5	95.0	90.0
Barley	80.6	62.5	70.4	84.6	55.0	66.7	81.8	67.5	74.0
Oats	63.2	40.0	49.0	78.6	36.7	50.0	70.6	40.0	51.1
Rapeseed	95.7	98.5	97.1	95.7	98.5	97.1	98.5	98.5	98.5
Canary Seed	63.2	66.7	64.9	88.9	44.4	59.3	71.4	55.6	62.5
Field Peas	92.5	94.2	93.3	94.3	96.2	95.2	94.0	90.4	92.2
Lentils	84.2	84.2	84.2	87.5	73.7	80.0	86.5	84.2	85.3
Flax	87.5	94.6	90.9	75.6	83.8	79.5	85.0	91.9	88.3
Grassland	91.4	93.7	92.5	89.2	93.7	91.4	86.7	91.1	88.9
Fallow	87.5	73.7	80.0	84.2	84.2	84.2	85.7	94.7	90.0
Classifier	SVM-POLY			ML					
Measure [%]	PA	UA	F	PA	UA	F			
Crop Type									
Wheat	84.4	93.8	88.8	81.6	74.5	77.9			
Barley	77.1	67.5	72.0	54.8	57.5	56.1			
Oats	61.1	36.7	45.8	38.1	26.7	31.4			
Rapeseed	99.3	97.8	98.5	97.0	96.3	96.7			
Canary Seed	76.9	55.6	64.5	28.1	50.0	36.0			
Field Peas	95.9	90.4	93.1	97.9	90.4	94.0			
Lentils	87.5	92.1	89.7	73.0	71.1	72.0			
Flax	91.9	91.9	91.9	69.0	78.4	73.4			
Grassland	86.9	92.4	89.6	91.0	89.9	90.4			
Fallow	87.5	94.7	90.0	64.3	94.7	76.6			

Table 8: Median classification accuracies for single crop types depending on classification algorithm used for the study area at Indian Head (CA). Two satellite images from June 2 and August 10 were used. PA: producer's accuracy; UA: user's accuracy; F: F-Measure.

After adding a second image from August 10th certain classes were classified much more precisely than with mono-temporal coverage only (cf. Table 8). However, the results of *Grassland* remained nearly stagnant or slightly lower with F-Measures between 86.6 and 90.8 %. In this configuration *Rapeseed* reached very accurate results with F-Values between 92 and 94.6 % for all classifiers. Other classes with F-Values over 80 % included *Wheat*, *Fallow* and *Field Peas.* In all cases the machine learning classifiers outperformed ML by up to 8 %. *Flax* and *Lentils* were classified at around 70 % accuracy. The remaining crop types *Barley*, *Oats* and *Canary Seed* were poorly classified due to a high level of confusion with W*heat*.

Classifier	ANN			RF			SVM-RBF		
Measure [%]	PA	UA	F	PA	UA	F	PA	UA	F
Crop Type									
Wheat	88.2	88.2	88.2	81.5	95.7	88.0	85.4	94.4	89.7
Barley	77.4	60.0	67.6	78.6	55.0	64.7	79.4	67.5	73.0
Oats	57.1	40.0	47.1	60.0	30.0	40.0	57.1	40.0	47.1
Rapeseed	96.4	99.3	97.8	98.5	97.8	98.2	97.8	99.3	98.5
Canary Seed	72.2	72.2	72.2	83.3	55.6	66.7	66.7	66.7	66.7
Field Peas	94.2	94.2	94.2	94.1	92.3	93.2	96.1	94.2	95.1
Lentils	82.5	86.8	84.6	91.2	81.6	86.1	91.2	81.6	86.1
Flax	83.3	94.6	88.6	79.5	94.6	86.4	89.5	91.9	90.7
Grassland	87.1	93.7	90.2	91.3	92.4	91.8	91.1	91.1	91.1
Fallow	85.0	89.5	87.2	81.8	94.7	87.8	89.5	89.5	89.5

Classifier	SVM-POLY			ML		
Measure [%]	PA	UA	F	PA	UA	F
Crop Type						
Wheat	83.5	94.4	88.6	81.0	73.9	77.3
Barley	77.8	70.0	73.7	50.0	50.0	50.0
Oats	73.3	36.7	48.9	34.5	33.3	33.9
Rapeseed	97.8	98.5	98.2	95.6	94.9	95.2
Canary Seed	84.6	61.1	71.0	34.8	44.4	39.0
Field Peas	94.2	94.2	94.2	95.9	90.4	93.1
Lentils	91.9	89.5	90.7	71.8	73.7	72.7
Flax	84.6	89.2	86.8	65.2	81.1	72.3
Grassland	91.3	92.4	91.8	92.1	88.6	90.3
Fallow	89.5	89.5	89.5	69.2	94.7	80.0

Table 9: Median classification accuracies for single crop types depending on classification algorithm used for the study area at Indian Head (CA). Three satellite images from June 2, August 10 and August 25 were used. PA: producer's accuracy; UA: user's accuracy; F: F-Measure.

Classifier	ANN			RF			SVM-RBF		
Measure [%]	PA	UA	F	PA	UA	F	PA	UA	F
Crop Type									
Wheat	81.3	91.9	86.3	79.7	95.0	86.7	80.7	93.8	86.8
Barley	66.7	50.0	57.1	70.8	42.5	53.1	78.1	62.5	69.4
Oats	35.7	16.7	22.7	40.0	6.7	11.4	25.0	10.0	14.3
Rapeseed	92.9	95.6	94.2	90.9	95.6	93.2	92.3	97.1	94.6
Canary Seed	66.7	66.7	66.7	81.8	50.0	62.1	75.0	50.0	60.0
Field Peas	86.5	86.5	86.5	88.9	92.3	90.6	86.5	86.5	86.5
Lentils	71.9	60.5	65.7	73.7	73.7	73.7	72.7	63.2	67.6
Flax	65.1	75.7	70.0	67.4	78.4	72.5	65.0	70.3	67.5
Grassland	88.5	87.3	87.9	88.1	93.7	90.8	88.8	89.9	89.3
Fallow	71.4	78.9	75.0	87.5	73.7	80.0	84.2	84.2	84.2

Classifier	SVM-POLY			ML		
Measure [%]	PA	UA	F	PA	UA	F
Crop Type						
Wheat	81.3	91.9	86.3	79.7	95.0	86.7
Barley	66.7	50.0	57.1	70.8	42.5	53.1
Oats	35.7	16.7	22.7	40.0	6.7	11.4
Rapeseed	92.9	95.6	94.2	90.9	95.6	93.2
Canary Seed	66.7	66.7	66.7	81.8	50.0	62.1
Field Peas	86.5	86.5	86.5	88.9	92.3	90.6
Lentils	71.9	60.5	65.7	73.7	73.7	73.7
Flax	65.1	75.7	70.0	67.4	78.4	72.5
Grassland	88.5	87.3	87.9	88.1	93.7	90.8
Fallow	71.4	78.9	75.0	87.5	73.7	80.0

Table 10: Median classification accuracies for single crop types depending on classification algorithm used for the study area at Indian Head (CA). Four satellite images from June 2, August 10, August 25 and September 5 and were used. PA: producer's accuracy; UA: user's accuracy; F: F-Measure.

Further improvements of the classification accuracies were observed with a third satellite coverage (cf. Table 9). The already well-recognised class *Rapeseed* achieved even stronger classification accuracies of more than 95 % in all measures and with all classifiers. With accuracies of more than 90 % *Field Peas* exhibited an increase of 5 to 12 % in F-Measure compared to only two satellite coverages, meanwhile *Grassland* did not show variation in classification accuracy between one, two or three image acquisitions.

The classes *Flax*, *Lentils*, *Fallow* and *Wheat* possessed a specific behaviour showing a strong dependency on the classifier used. Their F-Measures did not fall below 79.5 % while using machine learning classifiers, but only ranged from 72 to 78 % with ML. Moreover, both SVM-classifiers outperformed RF and ANN, in case of *Lentils* by up to 12.4 %, in case of *Wheat* by up to 3 %. The Cereals *Barley* and *Oats* showed improvements over using two images resulting in accuracies of around 70 % for *Barley*, but still only around 50 % for *Oats* with RF, ANN and both SVM-classifiers. The classification accuracy for *Canary Seed* remained constant at around 60 to 65 %.

The addition of the fourth and last satellite image from September 5th produced only minor improvements for *Canary Seed* and predominantly similar results for any other crop (cf. Table 10).

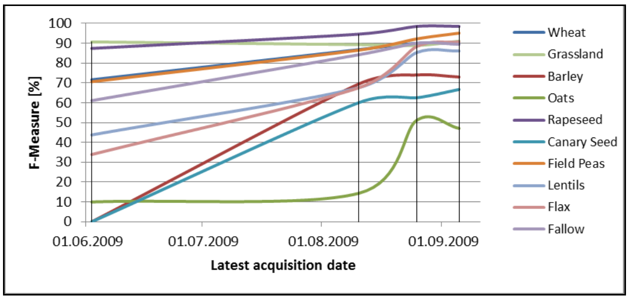

Figure 3: Crop specific classification results achieved with SVM-RBF in study area at Indian Head (CA) as a function of number of images used.

5. DISCUSSION

The classification accuracy varied strongly among the tested crop types. *Grassland* could be classified early in the growing season with only one satellite image. The addition of further satellite imagery did not increase its classification accuracy.

The best results were generally accomplished for *Rapeseed*, which can be accurately classified using two images, though the results still improved with a third coverage. Nearly as good results were achieved for *Field Peas*, which could be safely recognised in early August with two images. *Lentils*, *Flax* and *Fallow* achieved similar results over time. They reached maximum accuracies of around 90 % which is slightly lower than the accuracies of *Rapeseed* and *Field Peas*, but comparable to these of *Grassland*. The examined cereals, namely *Wheat*, *Barley* and *Oats*, exhibited different behaviour regarding their achieved accuracies. While the accuracies of *Wheat* were generally high, comparable to these of *Grassland*, *Barley* and especially *Oats* were classified much worse. Most misclassifications of the latter two classes could be attributed to false positive classifications of *Wheat*, which might have been a result of the skewed class sizes of the trainings sets. The role of relative and absolute size of a class used for training and its impact on the classification accuracy will have to be further investigated.

Classification results were strongly influenced by the number of satellite images used as well as the type of classifier. Both SVM classifiers outperformed RF and ANN in most cases. The poorest results by far were obtained with ML. The observed differences between ML and the used non-parametric classifiers are similar to the findings of other studies (Yang et al., 2011; Dixon & Candade, 2008; Huang et al., 2002). Nonetheless, in our study we also observed larger, in some cases even significant, differences among the machine-learning algorithms. These were found specifically on the early-season mono-temporal dataset, where overall accuracies diverged by up to 17 % between SVM-RBF and RF. With additional satellite imagery coverage the results converged to similar overall accuracies for all machine learning classifiers.

Classification accuracy improved with the number of images used. However, there seemed to be a threshold of maximum overall accuracy which could be achieved. It was just below 90 %. In September, by the time of the last image acquisition, some early sown crops were close to reaching maturity or were mature already. They had senesced and thus, no distinctive signal could be picked up by the sensors. Though, despite stagnant mean overall accuracies one specific classification behaviour was observed: While increasing the number of images, the classifier robustness increased strongly, which suggests less influence of the choice of training data and higher certainty of results in single classification runs. This development was even observed between three and four satellite images, which is in contrast to mean overall accuracies.

In the context of calculation complexity our results agree very well with these of other studies (Pal & Mather, 2003; Huang et al., 2002). ANN required by far the highest calculation times, whereas the training and testing of RF took usually longer than both SVM types. ML excelled in both stages with very short computation times. But all in all, *WEKA* performed the classifications very fast in a matter of seconds.

Artificial Neural Network produced good results, usually in between SVM and RF, but has many disadvantages. The complex architecture optimization, low calculation robustness and enormous training time outweigh the good mean classification accuracies. In this study, we found it to be the least favourable classifier among the machine-learning methods.

ML offers the most comfortable usage, where no parameters have to be optimized and calculation times are marginal. However, these advantages are outweighed by the poor classification performance.

Support Vector Machine with polynomial kernel (SVM-POLY) serves as an alternative to SVM-RBF with primarily non-significantly inferior classification results and slightly more complex setup.

Random Forest in contrast is easy to use, since only one variable needs to be set by the user. However, its classification accuracies when only one satellite coverage was used was the worst among machine-learning methods whereas its robustness was among the best.

After evaluating all measures, namely classification accuracy, robustness, calculation complexity and usability, Support Vector Machine with RBF-Kernel emerged as the best solution for the classification of different crop types using multi-temporal RapidEye data. This method excelled in classification performance and robustness and exhibited faster calculation time compared ANN.

REFERENCES

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R. J., Li, H., & Moran, M. S., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data [CD Rom]. In: *Proceedings of the Fifth International Conference on Precision Agriculture*, Bloomington, MN, USA, 16–19 July 2000.

Breiman, L., 2001. Random Forests, *Machine Learning* 45, pp. 5-32.

Chavez, P. S., 1996. Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering and Remote Sensing* 62(9), pp. 1025-1036.

Cortes, C. & Vapnik, V., 1995. Support-Vector Networks, *Machine Learning* 20(3), pp. 273-297.

Dash, J. & Curran, P., 2007. Evaluation of the MERIS terrestrial chlorophyll index (MTCI), *Advances in Space Research* 39(1), pp. 100-104.

Dixon, B. & Candade, N., 2008. Multispectral landuse classification using neural networks and support vector machines: one or the other, or both?, *International Journal of Remote Sensing* 29(4), pp. 1185-1206.

European Space Agency (ESA), 2009. *A glimpse of future GMES Sentinel-1 radar images*. [online] Available at: <http://www.esa.int/esaLP/SEM1FCANJTF_LPcampaigns_0.html> [Accessed 14 March 2012]

Haboudane, D.; Miller, J. R.; Pattey, E.; Zarco-Tejada, P. J. & Strachan, I. B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture, *Remote Sensing of Environment* 90(3), pp. 337-352.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, B. and Witten, I., 2009. The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Huang, C.; Davis, L. C. & Townshend, J. R. G, 2002. An assessment of support vector machines for land cover classification, *International Journal of Remote Sensing* 23, pp. 725-749.

Maas S.J., Rajan N., 2008. Estimating Ground Cover of Field Crops Using Medium-Resolution Multispectral Satellite Imagery. *Agronomy Journal* 100, pp. 320-327.

Pal, M & Mather, P. M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment* 86, pp. 554–565.

RapidEye, 2011. *Satellite imagery product specifications.* [online] Available at: http://www.RapidEye.de/upload/ RE_Product_Specifications_ENG.pdf [Accessed 01 Nov. 2011].

Rosenblatt, F., 1958. The Perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review* 65, pp. 386-408.

Rouse, J. W., Haas, R. H., Schell, J. A., and D. W. Deering, 1973. Monitoring Vegetation Systems in the Great Plains with ERTS. In: *Third ERTS Symposium*, Washington, DC, NASA SP-351, Vol. 1, pp. 309-317.

Rumelhart, D. E.; Hinton, G. E. & Williams, R. J., 1986. *Learning internal representations by error propagation*, MIT Press, Cambridge, MA, USA, pp. 318-362.

Yang, C.; Everitt, J. H. & Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification, *Computers and Electronics in Agriculture* 75, pp. 347-354.