

Land Cover Classification to Identify Wetlands Using Machine Learning

Perna Kapoor

Department of Resource Analysis, Saint Mary's University of Minnesota, Minneapolis, MN 55404

Keywords: Image Classification, Supervised Classification, GIS (Geographic Information Systems), Wetlands, CIR (Color-Infrared), North Dakota

Abstract

Wetlands provide a variety of ecological and economic functions that include water quality improvement, flood regulation and protection, groundwater recharge, shoreline stabilization, fish and wildlife habitat, agriculture production, aesthetics and biological productivity. Earlier traditional work on wetland extent assessment used relatively complex methods to compile wetland inventories (literature reviews, map interpretation, and digitizing), often giving incompatible and inconsistent results. Recent advances in sensor design and application have made remote sensing indispensable for wetland monitoring. For example, high resolution aerial imagery like National Agricultural Imagery Program (NAIP) and Color Infrared (CIR) imagery are used to distinguish wetlands. This study focuses on analyzing land cover classification to identify wetlands using machine learning. NAIP and CIR imagery were used to identify wetlands. After which, pixel-based supervised classifier methods were used as part of machine learning tools and a confusion matrix was calculated to find accuracy results of the three pixel-based supervised classifier methods. Overall, Random Trees Classifier had a substantially high level of accuracy compared to other classifier methods like Maximum Likelihood and Support Vector Machine. It means more number of wetlands were correctly identified using the Random Trees Classifier method. This study also helped to understand advantages of using machine learning, which is less time consuming in comparison to the manual process of identifying wetlands.

Introduction

Wetlands are valuable natural resources that provide many ecological services to both flora and fauna. Their benefits are a result of the natural hydrological and biogeochemical processes carried out in these ecosystems. These processes, which are sometimes called wetland functions, include hydraulic storage and recharge, bio-geochemical transformation, biomass production, and habitat (Marton, Creed, Lewis, Lane, Basu, Cohen, and Craft, 2015). In addition, these habitats are important forms of economic resources in

many countries in the form of recreation, fishing, waterfowl hunting, and animal grazing (Marton *et al.*, 2015). In recent times, wetlands have also become a popular topic in discussions of climate change because they contain 12% of the global carbon pool. Over time, wetlands have been drained, dredged, filled, leveled, and flooded to the extent that less than half of the original acreage remains (Dahl, 2000). Traditionally, wetlands are delineated using ground surveys. However, the surveys are difficult and time consuming (Yasouka, Yamagata, Tamura, Sugita, Pornprasertchai, Polngam,

Sripumin, Oguma, and Li, 1995). Remote sensing is one of the technologies that can provide cost and time-effective solutions to mitigate these problems (Goldberg, 1998). In addition, remote sensing technologies can supply the following information: (1) extent of wetlands, (2) identify the wetland resource as to type, (3) characterize the general wetland land cover type, (4) identify submergent and emergent wetlands.

Recognizing the importance of wetlands, Artificial Intelligence (AI) based extraction models have been recognized to discover patterns and images. Therefore, the goal of the current project is to analyze pixel-based supervised classifier methods to identify wetlands using machine learning and remote sensing in North Dakota. The specific goal was to investigate the accuracy levels and limitations of the automated methods of mapping wetlands and to discuss the implications of wetland mapping. Results of the study are intended to provide an evaluation of pixel-based supervised approaches to analyze an efficient way of wetland mapping.

Study Area

Study area selected for this project is peninsular land located to the south of Sanish and New Town of Mountrail County in North Dakota (Figure 1). This study area was selected because it has both glaciated and unglaciated land surfaces and helps in creating diverse set of training samples. Also, high resolution imagery was available for this area. Hence, it was a good site to test machine learning tool. There are more than 1 million wetland and lake basins in North Dakota, with densities of more than 10 wetlands per square mile in some areas (North Dakota Game and Fish Department, 2016). Lakes in North

Dakota are particularly susceptible to non-point source pollution, in part due to the great amount of agriculture in the state. Nearly all wetlands play a vital role in filtering clean water, storage of surface water, and crucial wildlife habitat.



Figure 1. Peninsular land to the south of Sanish and New Town city of North Dakota.

Data

The map image used in this research is a Color Infrared (CIR) image (Figure 2), which features recent high-resolution 1-meter aerial imagery of the study area. It is downloaded from ArcGIS online portal and is selected for classification because it is in 4 bands (Near Infrared, Red, Green and Blue bands) and gives a better indication of vegetation and presence of wetlands. CIR imagery is acquired at a one-meter Ground Sample Distance (GSD) with a horizontal accuracy that matches within six meters of photo-identifiable ground control points, which are used during image inspection. It is also called a “False color” image as it is not visible in this color to the human eye as any NAIP imagery is visible. The CIR image used was for the year 2019.



Figure 2. CIR image of study area.

Methods

This project focused on classifying land cover using Machine Learning to identify wetlands. To accomplish this, a pixel-based supervised classification approach was used. As per wetlands classification by National Wetland Inventory (NWI), the study area was divided into five different classes: wetland, open water body, farmland, service road/buildings, and vegetation. The training samples were distributed throughout the study area so that each class was represented in the training data. Fifty (50) samples were used for the study area per recommendations of protocol identified by GeoSpatial Services (2019).

The three methods that were used for machine learning consisted of Random Trees Classifier, Maximum Likelihood Classifier and Support Vector Machine (Blaschke and Lang, 2006). The classification process and testing of classifier methods were conducted in ArcGIS Pro (10.4) software. The results of accuracy were calculated using accuracy assessment and calculating the kappa coefficient.

Random Trees Classifier

Random Trees is a collection of individual decision trees where each tree is generated from different samples and subsets of the

training data. The idea behind calling these Decision Trees (DT) is that for every pixel that is classified, several decisions are made in rank order of importance. The decision trees are known to produce results of higher accuracies in comparison to traditional approaches such as the box and minimum distance to means classifiers, but the performance of decision trees can be affected by a number of factors including pruning and boosting methods used and decision thresholds (Mahesh and Mather, 2003).

This method is called Random Trees because each dataset is classified a number of times based on a random sub selection of training pixels, thus resulting in many decision trees. To make a final decision, each tree has a vote. This process works to mitigate overfitting. Random Trees is a supervised machine-learning classifier based on constructing a multitude of decision trees, choosing random subsets of variables for each tree, and using the most frequent tree output as the overall classification. The following algorithm is used for this classifier method:

$$\sum_i^n a(i)x(i) \leq c$$

for multivariate decision trees or simply $x(i) > c$ for univariate decision trees. Where $x(i)$ represents the measurement vectors on the n selected features and a is a vector of linear discriminate coefficients while c is the decision threshold (Brodley and Utgoff, 1992). The DTs are known to produce results of higher accuracies in comparison to traditional approaches such as the “box” and “minimum distance to means” classifiers.

Maximum Likelihood Classifier

Maximum Likelihood Classification

(MLC) assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. Each pixel is assigned to the class that has the highest probability (that is, the maximum likelihood). The tool considers both the means and covariance of the class signatures when assigning each cell to one of the classes represented in the signature file. When the default equal option for probability weighting is specified, each cell is assigned to the class to which it has the highest likelihood of being a member. The advantage of the MLC as a parametric classifier is that it takes into account the variance–covariance within the class distributions and for normally distributed data (Erdas, 1999). The following algorithm is used for this classifier method:

$$D = \ln(a_c) - [0.5 \ln(|\text{cov}_c|)] - [0.5 (X - M_c)^T (\text{cov}_c^{-1}) (X - M_c)]$$

Where weighted distance or likelihood D of unknown measurement vector X belonging to one of the known classes in the equation and M_c is maximum likelihood classifier based on the Bayesian equation.

Support Vector Machine Classifier

The support vector machines (SVMs) are a set of related learning algorithms used for classification and regression. Like the decision trees classifiers, the SVM are also non-parametric classifiers. In other words, it does not require data to fit normal distribution. The theory of the SVM was originally proposed by Vapnik and Chervonenkis (1971) and later discussed in detail by Vapnik (1999). The objective of a support vector classifier is to find a hyperplane in an N dimension (N - number of features) that distinctly classifies data

points. Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane.

The easiest way to train the SVM is by using the linearly separable classes algorithm:

$$\begin{aligned} WX_i + b &\geq +1 \text{ for all } y = +1. \\ WX_i + b &\leq -1 \text{ for all } y = -1. \end{aligned}$$

According to Osuna and Freud (1997) if the training data with k number of samples is represented as $\{X_i, Y_i\}$, $i = 1, \dots, k$ where $X = RN$ is an N -dimensional space and $y = \{-1, +1\}$ is a class label then these classes are considered linearly separable if there exists a vector W perpendicular to the linear hyper-plane (which determines the direction of the discriminating plane) and a scalar b showing the offset of the discriminating hyper-plane from the origin. For the two classes, it means class 1 represented as -1 and class 2 represented as $+1$, two hyper-planes can be used to discriminate the data points in the respective classes.

Accuracy Assessment

After the land cover classification was done and different classifier methods were tested, accuracy results were calculated. This was done with the help of a confusion matrix available in the ArcGIS Pro environment. Accuracy assessment was used to quantify how well the classified results represent the underlying raster data and evaluate the accuracy of each class to determine the overall quality of the image classification (Foody, 2002).

To see the assessment results, equalized stratified random points were generated (Twenty for each class) using an Accuracy Assessment tool and two new fields (classified and ground truth) were

added in the attribute table. Depending on the class value assigned to each point, the ground truth values were changed looking at the location in the image. After this, the confusion matrix was run to evaluate the kappa coefficient. This process was repeated for all three classifier methods and hence overall accuracy results of the methods were compared.

The confusion matrix indicates the classification results for each of the random points. The confusion matrix gave a kappa statistic which explained an overall accuracy for each of the classifier methods. Kappa coefficients range from 0 to 1. It is considered a substantial agreement if the Kappa coefficient falls between 0.61 to 0.8, and almost perfect agreement if the Kappa coefficient falls between 0.81 to 0.99 (Landis and Koch, 1977). In this study, substantial agreement was selected as the standard for accuracy of the classification method results as shown in Table 1.

Table 1. Accuracy standard for kappa coefficient calculated for classifier methods.

Kappa Coefficient	Accuracy Standard
<0	No Agreement
0 - .20	Slight
.21 - .40	Fair
.41 - .60	Moderate
.61 - .80	Substantial
.81 - 1.0	Perfect

The formula for the kappa coefficient (κ) is:

$$k = \frac{Po - Pe}{1 - Pe} = 1 - \frac{1 - Po}{1 - Pe}$$

where P_o is the relative observed agreement among raters (identical to accuracy), and P_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities

of each observer randomly seeing each category. If the raters are in complete agreement, then $k = 1$. If there is no agreement among the raters other than what would be expected by chance (as given by P_e), $k = 0$.

Confusion Matrix Analysis

This matrix explains errors of omission and commission and derives a kappa index of agreement and an overall accuracy between the classified map and the reference data (Geneletti and Gorte, 2003). In the matrix table components U_Accuracy stands for user's accuracy. It represents the fraction of pixels classified correctly per total classifications. It also shows false positives, where pixels are incorrectly classified as a known class when they should have been classified as something else. User's accuracy is also referred to as errors of commission, or type 1 error. The data to compute this error rate is read from the rows of the table. P_Accuracy stands for producer's accuracy and represents the fraction of pixels classified correctly per total ground truths. It also shows false negatives, where pixels of a known class are classified as something other than that class. Producer's accuracy is also referred to as errors of omission, or type 2 error. The data to compute this error rate is read in the columns of the table. Accuracy is represented from 0 - 1, with 1 being 100 percent accurate.

Results

Confusion Matrix Results

The accuracy result for the three classifier methods are shown in Table 2, Table 3, and Table 4. Comparison of results along with run time for each classifier method is

shown in Table 5. Figures 3, 4, and 5 show the classified raster images for the supervised machine learning methods.

Following is the abbreviation used in confusion matrix for different classes:

- C_10 = Wetland
- C_20 = Open Water Body
- C_30 = Farmland
- C_40 = Service Road/Buildings
- C_50 = Vegetation

Table 2. Confusion Matrix for Random Tree Classification. Data corresponding to class value here is used to calculate User Accuracy (U_Accuracy which is Fraction of pixels classified correctly per total classifications), Producer Accuracy (P_Accuracy, Fraction of pixels classified correctly per total ground truths) and Kappa coefficient (Kappa, Overall assessment of the accuracy of the classification).

ClassValue	C_10	C_20	C_30	C_40	C_50	Total	U_Accuracy	Kappa
C_10	8	0	4	1	7	20	0.4	0
C_20	0	20	0	0	0	20	1	0
C_30	0	0	17	0	3	20	0.85	0
C_40	1	0	3	13	3	20	0.65	0
C_50	1	0	0	0	19	20	0.95	0
Total	10	20	24	14	32	100	0	0
P_Accuracy	0.8	1	0.708333	0.928571	0.59375	0	0.77	0
Kappa	0	0	0	0	0	0	0	0.7125

Table 3. Confusion Matrix for Maximum Likelihood Classification. Data corresponding to class value here is used to calculate User Accuracy (U_Accuracy which is Fraction of pixels classified correctly per total classifications), Producer Accuracy (P_Accuracy, Fraction of pixels classified correctly per total ground truths) and Kappa coefficient (Kappa, Overall assessment of the accuracy of the classification).

ClassValue	C_10	C_20	C_30	C_40	C_50	Total	U_Accuracy	Kappa
C_10	6	2	0	1	11	20	0.3	0
C_20	0	20	0	0	0	20	1	0
C_30	2	0	15	0	3	20	0.75	0
C_40	2	0	6	6	6	20	0.3	0
C_50	1	0	1	0	18	20	0.9	0
Total	11	22	22	7	38	100	0	0
P_Accuracy	0.545455	0.909091	0.681818	0.857143	0.473684	0	0.65	0
Kappa	0	0	0	0	0	0	0	0.5625

Table 4. Confusion Matrix for Support Vector Machine Classification. Data corresponding to class value here is used to calculate User Accuracy (U_Accuracy which is Fraction of pixels classified correctly per total classifications), Producer Accuracy (P_Accuracy, Fraction of pixels classified correctly per total ground truths) and Kappa coefficient (Kappa, Overall assessment of the accuracy of the classification).

ClassValue	C_10	C_20	C_30	C_40	C_50	Total	U_Accuracy	Kappa
C_10	5	0	5	1	9	20	0.25	0
C_20	0	20	0	0	0	20	1	0
C_30	1	0	15	0	4	20	0.75	0
C_40	2	2	1	12	3	20	0.6	0
C_50	1	0	0	0	19	20	0.95	0
Total	9	22	21	13	35	100	0	0
P_Accuracy	0.555556	0.909091	0.714286	0.923077	0.542857	0	0.71	0
Kappa	0	0	0	0	0	0	0	0.6375

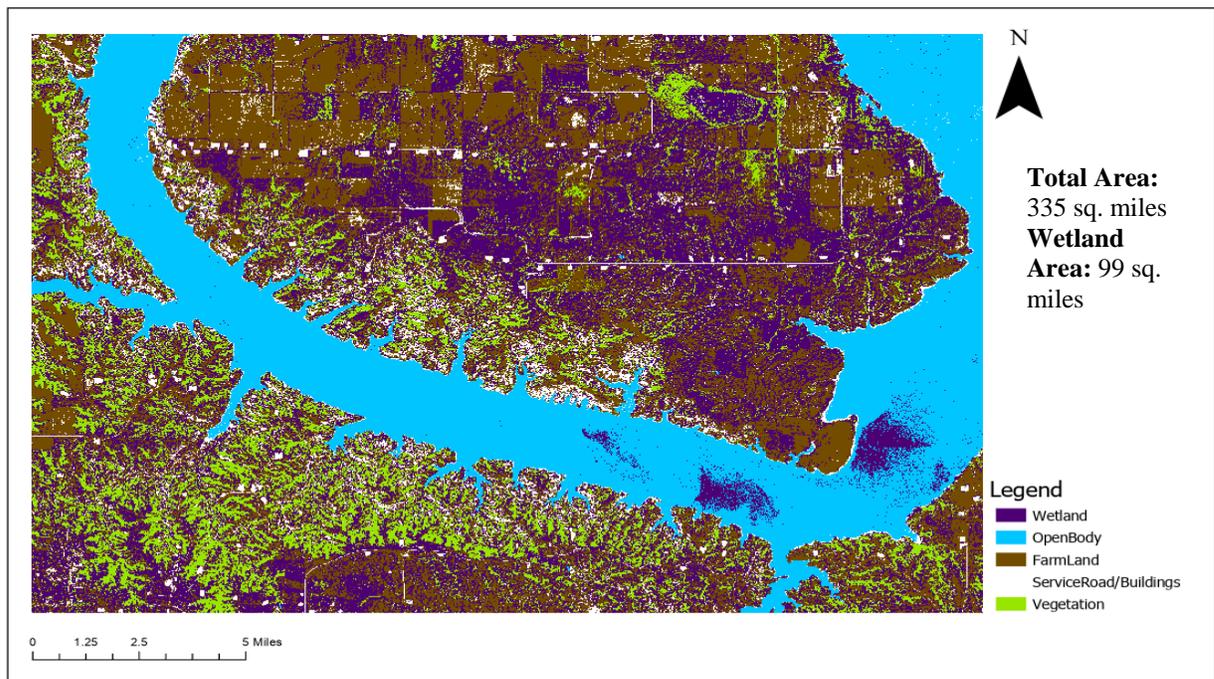


Figure 3. Classified Raster using Random Trees Classification.



Figure 4. Classified Raster using Maximum likelihood Classification.

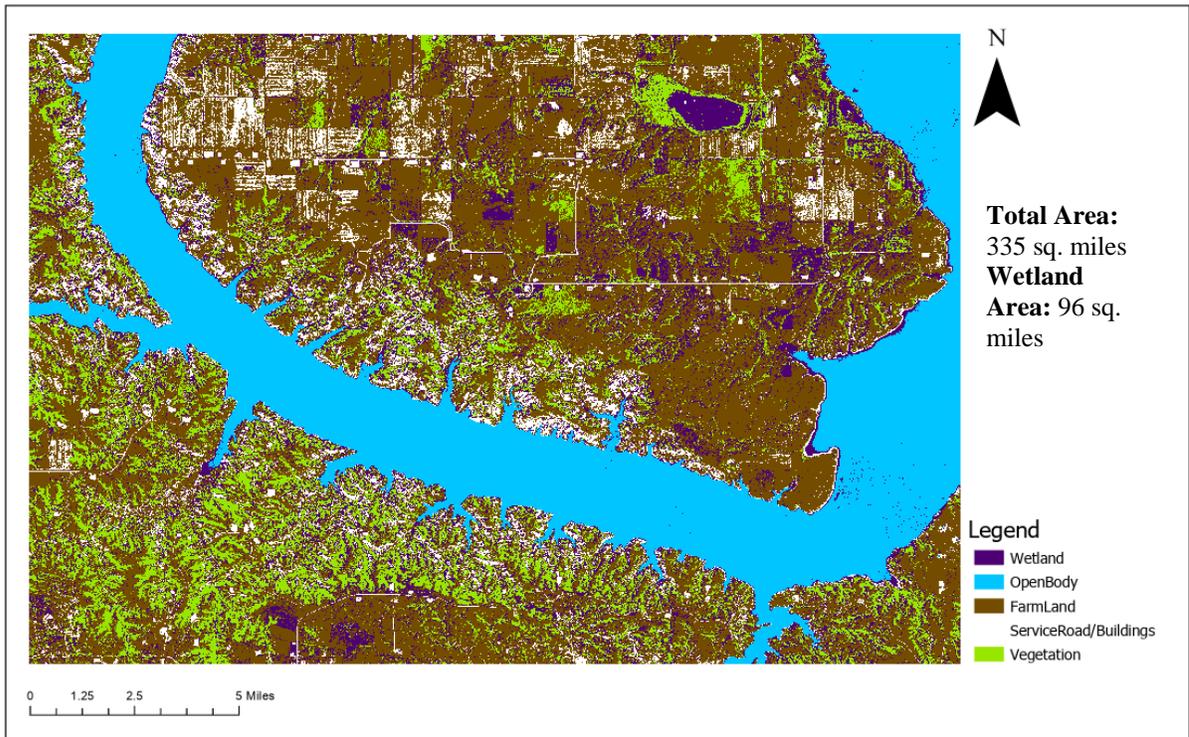


Figure 5. Classified Raster using Support Vector Result.

Table 5. Comparison of result for machine learning methods along with their respective run time.

Machine Learning Methods	Kappa Coefficient	Study Area (per Sq. Mile)	Run Time For each classifier (Hours)
Random Trees Classifier	0.71	335	7
Maximum Likelihood Classifier	0.56	335	6
Support Vector Classifier	0.63	335	8

Discussion

This study was conducted for land cover classification, so the study area was divided into five different categories and pixel-based supervised classifier methods were run to identify wetlands using machine learning tools. Each classifier method ran its own algorithm and gave an output by classifying land cover based on the training samples provided. The results were interesting and the difference in land cover classification was visible in each output. Random Trees classifier (Figure 3) presented the output by averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance. In this case, most of the wetlands were correctly identified, and the kappa coefficient was (0.71) which was significant as explained in (Table 1). Even other classes were mostly identified at the correct locations with very few errors. Maximum Likelihood Classifier (Figure 4) presents the output by assigning each pixel to a class that has the highest probability, that is, the maximum likelihood of belonging to that category. In this case the kappa coefficient was (0.56) which was not significant as explained in (Table 1) and most of the land categories were misidentified by the tool. It means wetlands and other class categories were not correctly identified. The results also show that class categories and ground truth did not match each other at many locations. Support Vector classifier (Figure 5) presents the output by

performing the classification by finding the hyper-plane that differentiates all the classes very well. In this case the kappa coefficient was (0.63) which was also significant as shown in (Table 1) and it could be said that some wetlands were correctly identified but other land categories were misidentified at some places like some of the farmland area was confused with wetland area and vegetation was also identified at some wrong areas. The results were such because it was possible that supervised training method would have misidentified pixels while classifying the classes which was a disadvantage of using pixel-based classification.

Hence, amongst the three classifier methods it could be said that results for Random Trees was acceptable as it had the highest kappa coefficient amongst the three classifier methods (Table 5). Also, it correctly identified maximum wetland area amongst the three classifier methods (Figure 3). The results were also compared against the manual classification process of identifying wetlands (Figure 6). The same study area could take approximate ninety (90) hours to digitize manually and identify wetlands (Table 6) in comparison to machine learning process which took twenty-one (21) hours in total to identify wetlands.

Few limitations were experienced in this research. For example, samples decided for training were at editor's discretion. It is possible representative samples were not chosen and may be some

wetland samples were left out.

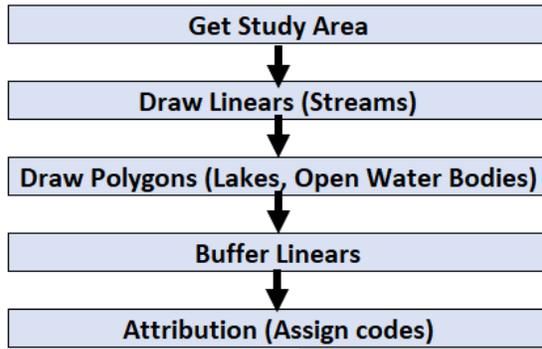


Figure 6. Manual Classification process.

Table 6. Results for manual vs machine learning in terms of time taken to identify wetlands.

Classification Method	Approximate time taken (in hours)
Random Trees Classifier	7
Maximum Likelihood Classifier	6
Support Vector Classifier	8
Manual (Refer to Figure 6)	90

Second, to calculate accuracy, random points were generated on the entire study area. It was difficult to say that the random points generated by the tool were the most representative to calculate accuracy. In addition, if points were manually selected then the results could have varied significantly. Lastly, another limitation was not to be able to use unsupervised methods of machine learning along with pixel-based supervised methods. It is because unsupervised methods classified the image in many categories, but it was difficult to associate the resulting classes with Land Cover categories. Hence, the results for identifying wetlands could not be established for supervised vs. unsupervised method.

Conclusions

This project analyzed land cover classification to identify wetlands using machine learning. The study area was divided into five different classes: wetland, open body, farmland, service road/buildings, and vegetation. By using pixel-based supervised classification, wetlands were identified. Results of the three classifier methods used were compared as their accuracy was calculated using a confusion matrix and the kappa coefficient. Random Trees and Support Vector Machine gave substantial levels of accuracy followed by the maximum likelihood classifier method where the level of accuracy was not substantial. This study may be helpful to organizations working on wetlands conservation as machine learning reduces time and effort otherwise would be needed to manually digitize wetlands (Table 6). It can also be useful to decision makers and government organizations who want to design quick solutions for wetland preservation.

Acknowledgements

I would like to thank John Ebert and Greta Poser for making GIS/DIGA program a fun experience to learn. It has added a new dimension to my career prospects ahead. I would also like to thank Andy Robertson, Kevin Benck and Alexander Rabine at Saint Mary's Geospatial Services Office at Minneapolis for letting me part of wetland conservation Fort Berthold Reservation, North Dakota project and for walking me through wetland digitizing process. Lastly, I would thank my husband for his constant support and encouragement.

References

Blaschke, T., and Lang, S. 2006.

- Object based analysis for automated information extraction-a synthesis MAPPS/ASPRS Fall Conference, San Antonio, Tx, USA.
- Brodley, C.E., and Utgoff, P.E. 1992. Multivariate versus univariate decision trees. Technical Report 92-8. University of Massachusetts, Amherst, MA, USA.
- Dahl, T.E. 2000. *Status and trends of wetlands in the conterminous United States 1986 to 1997*, U.S. Dept of the Interior, Fish and Wildlife Service, Washington D.C., 82 p.
- Erdas Inc. 1999. Erdas Field Guide Erdas Inc., Atlanta, Georgia.
- Foody, G.M. 2002. Status of land cover classification accuracy assessment *Remote Sensing of the Environment*, 80, pp. 185-201.
- Geneletti, D., and Gorte, B.G.H. 2003. A method for object-oriented land cover classification combining Landsat TM and aerial photographs. *International Journal of Remote Sensing*, 24 pp. 1273-1286.
- GeoSpatial Services. 2019. <http://www.geospatialservices.org/>.
- Goldberg, J. 1998. Remote Sensing of Wetlands: Procedures and Considerations. Retrieved January 24, 2003 from <http://www.vims.edu/rmap>.
- Landis, J.R., and Koch, G.G. 1977. The Measurement of Observer Agreement for Categorical Data. *Journal of Biometrics*, 33, 159-174.
- Mahesh, P., and Mather, P.M. 2003. An assessment of the effectiveness of the decision tree method for land cover classification *Remote Sensing of the Environment*, 86, pp. 554-565.
- Marton, J.M., Creed, I.F., Lewis, D.B., Lane, C.R., Basu, N.B., Cohen, M.J., and Craft, C.B. 2015. Geographically isolated wetlands are important biogeochemical reactors on the landscape. *Bioscience* 65 (4), 408–418.
- National Wetlands Inventory. 2018. US Fish and Wildlife Service (USFWS). 2018 Retrieved July 11, 2018 from <https://www.fws.gov/wetlands/>.
- North Dakota Game and Fish Department. 2016. <https://gf.nd.gov/>.
- Osuna, E.E., and Freud, R. 1997. *Support Vector Machines: Training and Applications*, A.I. Memo No. 1602, C.B.C.L. Paper No. 144 Massachusetts Institute of Technology and Artificial Intelligence Laboratory, Massachusetts.
- Vapnik, W.N., and Chervonenkis., A.Y. 1971. On the uniform convergence of the relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 17, pp. 264-280.
- Yasouka, Y., Yamagata, Y., Tamura, M., Sugita, M., Pornprasertchai, J., Polngam, S., Sripumin, R., Oguma, H., and Li, X. 1995. *Monitoring of Wetland Vegetation Distribution and Change by using Microwave Sensor Data*, Proceedings of ACRS, Retrieved November 29, 2003 from <http://www.gisdevelopment.net/aars/acrs/1995/ts1/ts1001pf.htm>.