

Window Histogram for Characterizing Basins

Shuhang Li & Anne Denton



Department of Computer Science, North Dakota State University



INTRODUCTION

In this research, our goal is to identify the relationship between vegetation density and low-lying areas. We use a window-based approach to preprocess the data within Fair Mount area. The distribution of NDVI and near-infrared data inside each window will be summarized using histograms. Each column inside a histogram will then be treated as a separate attribute. Using the preprocessed data set, we build a decision tree classifier that has a sensitivity of 67% and a specificity of 81%.

OBJECTIVES

- Use window-based approach to preprocess NDVI and near-infrared data within Fair Mount area.
- Build a decision tree model to classify the preprocessed data sets.
 1. The whole map was cut into two parts. The larger upper part is used as training data set and the smaller lower part is used as testing data set. The basins of the whole map were identified and their coordinates were given by Dr. Xuefeng Chu, NDSU.
 2. The text files are processed into window-based data sets using the algorithm in a particular format to be classified using the classifier.
 3. The attributes are the summarized NDVI and near-infrared data. Each observation includes the above data inside a single window.
- Identify the relationship between low-lying areas and the attributes based on the built model.

METHODS

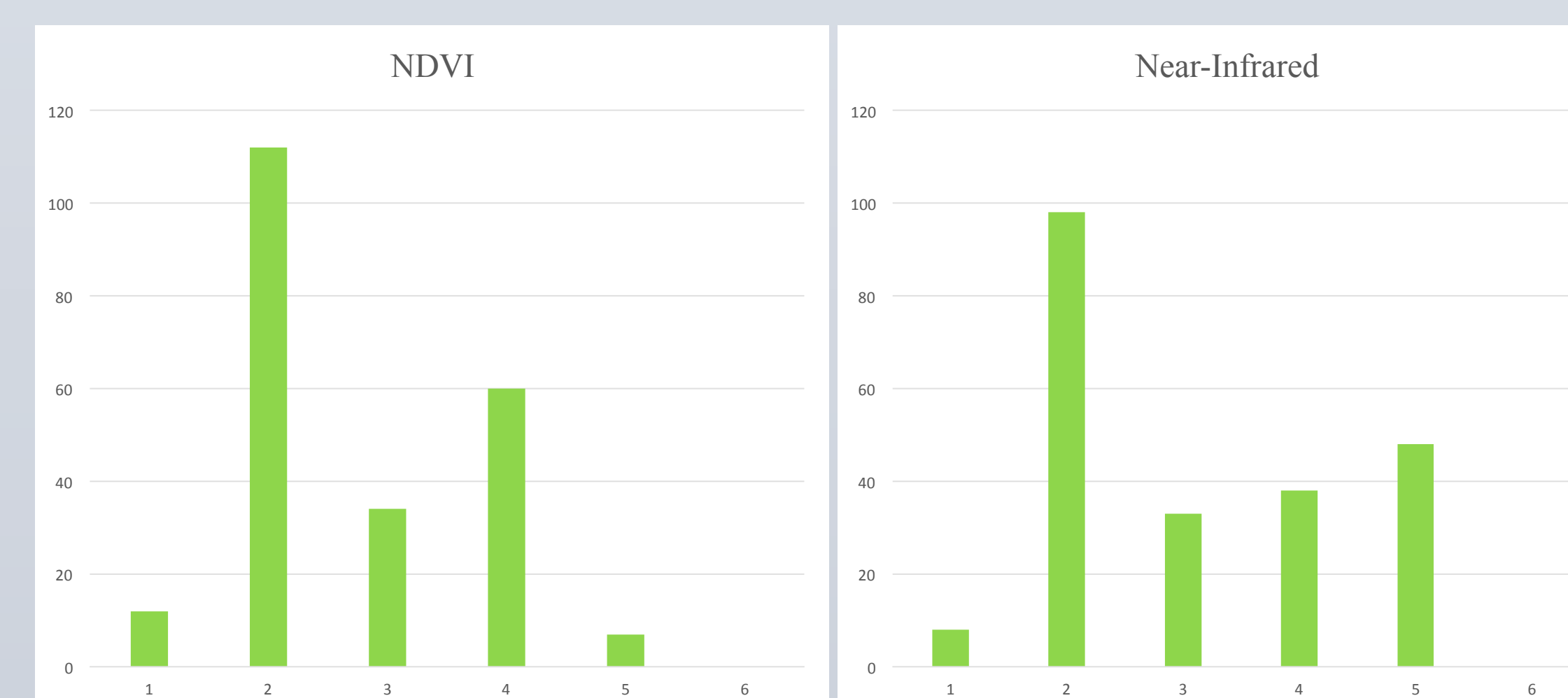
Process:

1. Read the input text files, which contain the ndvi and near infrared information, to gather needed data for identifying features.
2. Distribute the data into different windows and make histogram-based data format for each window.
3. Build a decision tree using the data sets that are produced in Step 2.
4. Print out the decision tree model and the information gain for each attribute while building the model.
5. Reduce size of the tree by selecting 50% of the attributes that has the highest information gain. Build a new model based on these attributes.

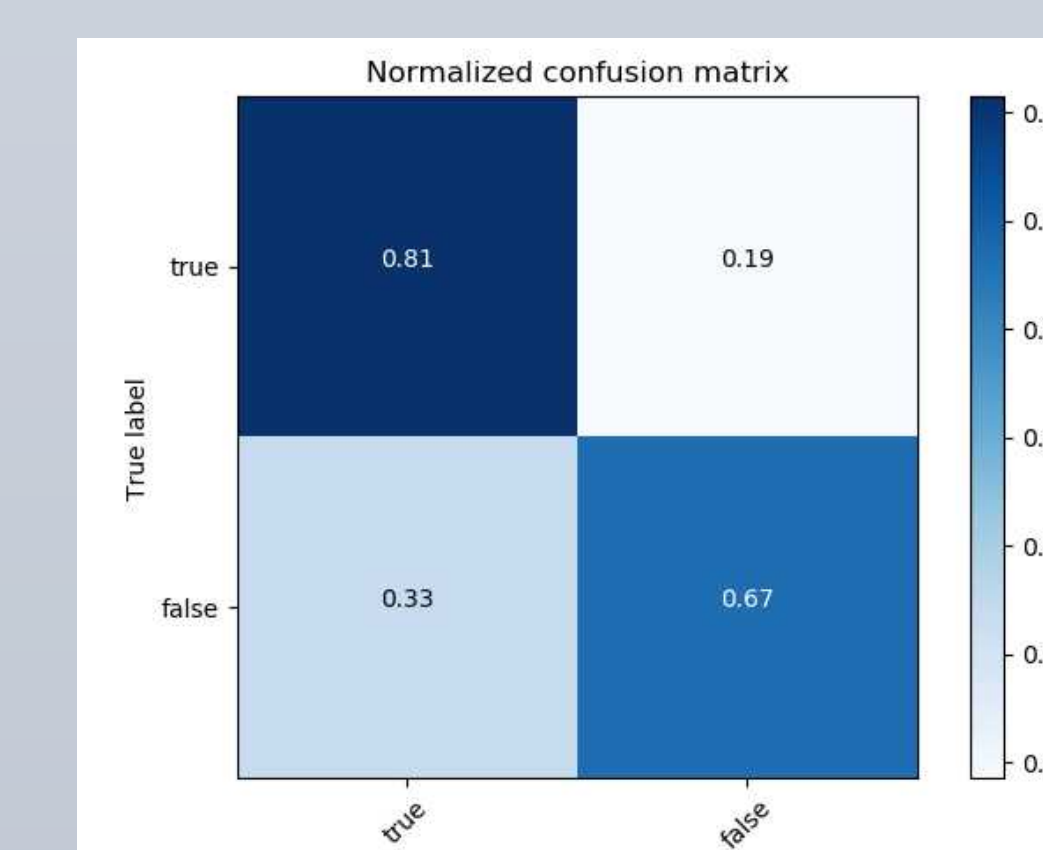
Part of text based raw data:

```
K,Y,Z
682455,5099765,-0.0415536501
682465,5099765,-0.0233036336
682475,5099765,-0.0237412546
682485,5099765,-0.0360228941
682495,5099765,-0.0211235955
682505,5099765,-0.027578475
682515,5099765,-0.018413946
682525,5099765,-0.0364997648
682535,5099765,-0.0334884785
682545,5099765,-0.0120817842
682555,5099765,-0.0397611931
682565,5099765,-0.0364400856
682575,5099765,-0.0510603003
682585,5099765,-0.0317279175
682595,5099765,-0.0263861436
682605,5099765,-0.0134630613
682615,5099765,-0.0183150191
```

NDVI and Near-Infrared histograms after preprocessing:



Classification result using decision tree:



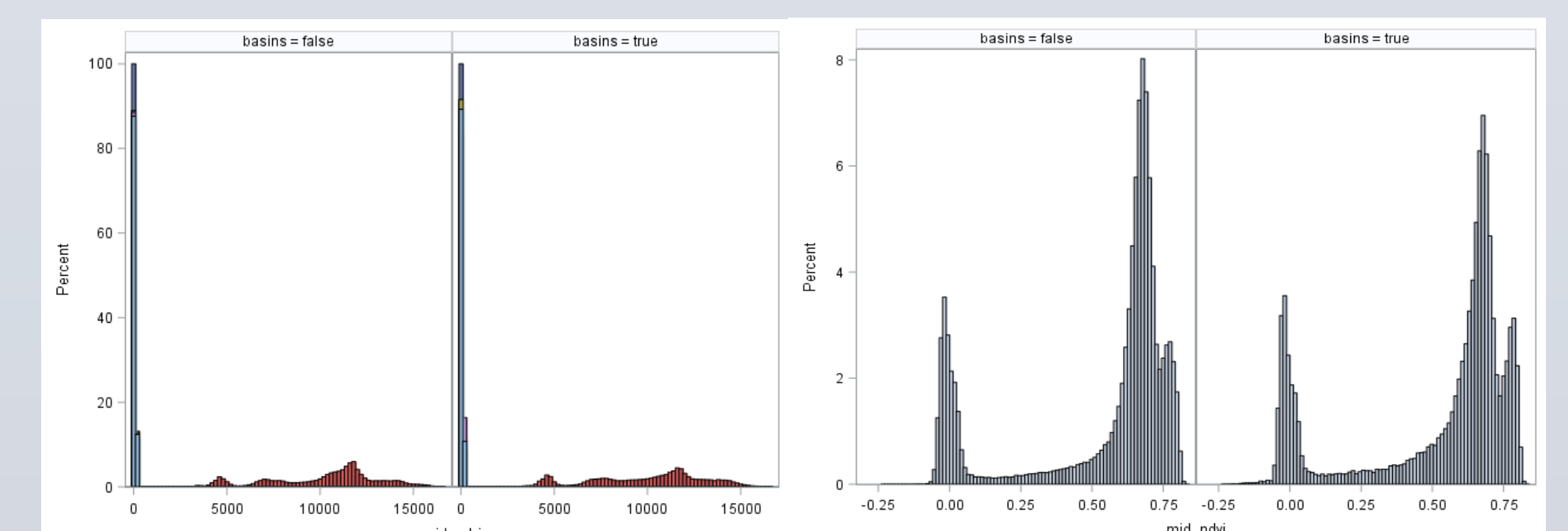
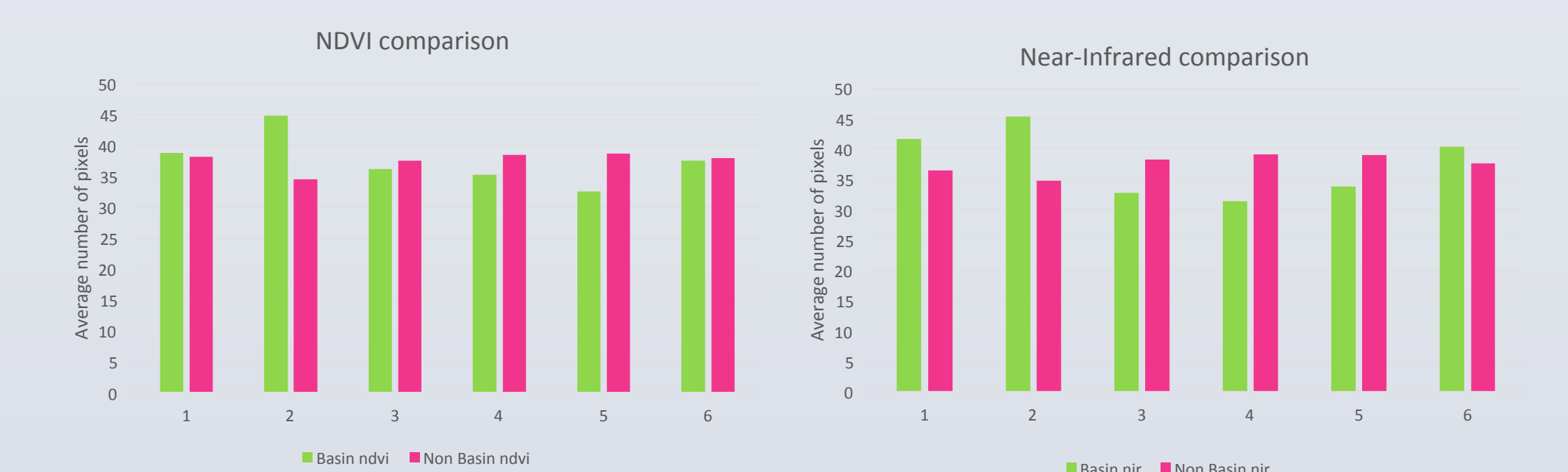
RESULTS

Training dataset: Upper section of Fairmount site

Testing dataset: Lower section of Fairmount site

Window Size: 15 * 15

From the comparison based on NDVI and Near-Infrared data from training data sets in the pictures below we can see that the area that contains basins has less vegetation density. Although the difference is not significant, both data support this theory.



Acknowledgments

This material is based upon work supported by the National Science Foundation through grant IIA-1355466

Contact

Shuhang Li: (218)275-3458, shuhang.li@ndsu.edu
Dr. Anne Denton: anne.denton@ndsu.edu