

A Machine Learning Framework for Detecting Land Slides On Earthen Levees Using Spaceborne SAR Imagery

Majid Mahrooghy^{1,*}, James Aanstoos¹, Saurabh Prasad², Khaled Hasan¹,
and Nicolas H. Younan³

¹Geosystems Research Institute, Mississippi State University
Box 9627, Mississippi State, MS 39762

²Department of Electrical and Computer Engineering, University of Houston,
Houston, Texas 77004, USA

³Department of Electrical and Computer Engineering, Mississippi State University,
Mississippi State, MS 39762, USA

Submitted to the JSTARS Special Issue on

Eo Approaches For Large Area Land Monitoring With Multiple Sensors And Resolutions

July 30, 2012

* Corresponding Author: Majid Mahrooghy, Geosystems Research Institute,
Box 9627, Mississippi State University, Mississippi State, MS 39762.
E-mail: majid@gri.msstate.edu Phone: 662-325-6927

ABSTRACT

Earthen levees have a significant role in protecting large areas of inhabited and cultivated land in the US from flooding. Failure of the levees can result in loss of life and property. Slough slides are among the problems which can lead to complete levee failure during a high water event. In this paper, we develop a method to detect such slides using X-band SAR data. Our proposed methodology includes; 1) Radiometric normalization of the TerraSAR image using Digital Elevation Map (DEM) data, 2) extraction of features including backscatter and texture features from the levee, 3) training a support vector machine classifier; and 4) testing on the area of interest. Ground-truth data are collected from slides and healthy areas of the levee. The study area is part of the levee system along the lower Mississippi River in the United States. The output classification shows the two classes of healthy and slide areas. The results show classification accuracies of approximately 92% for detecting the slide pixels when more than 100 training samples are used.

Index Terms - Hazards, synthetic aperture radar, feature extraction, support vector machine

1. INTRODUCTION

There are more than 150,000 kilometers of levee structure with different designs and conditions over the entire US which have a significant role in protecting large areas of populated and cultivated land from flooding. Since a failure of the levees can threaten the loss of life and property, levee system monitoring to detect failures and damage is important. Such detection using algorithms and techniques based on remote sensing images can help levee managers and federal agencies to identify vulnerable levee sections and repair them rapidly with lower costs [1]. Using remote sensing imagery is a more efficient and cost effective way to detect failures than traditional methods of frequent site visits. Different types of damage occur to the levee system that can trigger a complete failure especially during a high water event [2]. Slough slides occurring along levees leave sections vulnerable to seepage and failure during high water events. Hossain *et al.* [3] have discussed the cause of slides over a levee. Since the roughness and the texture of a slide are different from the healthy area, the corresponding radar backscatter is different, and therefore can be used to distinguish between slide and healthy pixels. Any early detection of slide events can assist levee managers to find and fix them and prevent costlier damage in the future. In addition to roughness and texture, other features such as soil moisture patterns and differences in the type of vegetation (over a slide area and the surrounding) can be used to detect the slide [3], [4]. Soil moisture monitoring can be an effective way to identify any leakage or anomaly occurring along a levee system [4].

Synthetic Aperture Radar (SAR) is able to provide valuable information which can be used in quantitative monitoring of the earth's surface [5]. Polarimetric SAR has

been recently utilized for many applications in order to measure various geophysical parameters. Most studies on polarimetric SAR have been carried out to extract information on the earth's surface characteristics such as surface roughness and soil moisture content in bare soil and vegetated areas. In soil moisture assessment, different regression models based on the relationship between backscatter coefficients and soil moisture content have been developed [6], [7].

In this paper, an algorithm based on TerraSAR-X [8] data is developed to detect landslides on levees. The paper is organized as follows: The data used in this study is explained in section 2. In section 3, the method and a block diagram of the algorithm are described. The results of applying the algorithm are discussed in the results and validation section, and finally a conclusion and summary of this study are provided in section 5.

2. DATA SET

Our area of study is part of the levee system along the lower Mississippi River and the western boundary of the state of Mississippi. Radar imagery from the German TerraSAR-X satellite was acquired over the study area. TerraSAR-X is a SAR sensor imaging at 9.65 GHz with a variable incidence angle and ground resolution. It acquires images in stripmap (SM), high resolution spotlight (HS), spotlight (SL), and scanSAR (SC) modes [8]. In this research, we use dual polarized HH/VV images from the HS mode acquired on September 15, 2010 with a 1.5 m spatial resolution. The images were ellipsoid corrected (EEC) and spatially enhanced detected products provided by the German Space Agency (DLR©2010). A satellite such as TerraSAR-X has an advantage

for monitoring levees versus an airborne platform since it has a better temporal resolution with a lower cost of data acquisitions.

We also used a 1.5 m per pixel DEM produced from Lidar data collected in 2009-2010 at a 1 m ground resolution. This DEM helps in obtaining the local incidence angle which is used for calibration of the raw data.

Fig. 1(a) shows an example of a typical levee slough slide. Fig. 1(b) shows a buffered HH-SAR image of the levee area where a slide occurred. The corresponding brightness of the slide is higher than the healthy area. This high brightness results from the greater surface roughness over the slide area.

3. METHODOLOGY

In this study we are developing a methodology to detect slides from background using a Support Vector Machine (SVM) and TerraSAR-X data. A block diagram of the methodology is depicted in Fig. 2. First the Single Look Slant Complex (SSC) TerraSAR-X images product [9] representing the reflectivity from the earth's surface is calibrated and normalized using the following formula [9], [10], [11].

$$\sigma_0 = (k_s \cdot |DN|^2 - NEBN) \cdot \sin \theta_{loc} \quad (1)$$

In this equation, k_s is the calibration and processor scaling factor. NEBN is the Noise Equivalent Beta naught which provides the influence of different noise contributions to the signal. Both k_s and NEBN are given in the TerraSARX product. DN is the digital number giving the value of a pixel in the products. θ_{loc} is the local incidence angle which is the angle between the radar beam and a line that is normal to the surface obtained by DEM. Since we are examining the slide detection only across the levee, after the radiometric normalization of TerraSAR images, a 30m wide study area on both sides

of the levee center line is extracted from the full SAR scene for further processing. In the next step, features are extracted from this area. These include radiometric features of HH, VV, HH/VV and texture features. The texture features include wavelet [12] and statistical features which are obtained by using a sliding window with length 7. The wavelet features are the mean and standard deviation (STD) of the coefficients in two levels of decomposition. Each of the decomposition levels includes the horizontal, vertical, and diagonal detail coefficients. Therefore, we have one approximate and 7 detail coefficients sections. The wavelet features are the mean and STD of each coefficient sections. Thus, 14 wavelet features are obtained for each pixel. Note that the 'db2' wavelet family is used as the wavelet filter for this work. Statistical features are the mean and STD of the sliding window for each pixel. Therefore, for every band HH and VV we have 2 radiometric and 16 texture features (2 statistics and 14 wavelet features). Hence, the total number of features including radiometric features (HH, VV, and HH/VV) is 35. Fig. 3 shows some features computed over the levee buffer. Fig. 3(a) shows the VV-STD feature for each pixel which is the mean of STD over a sliding window centred in the pixel for VV polarization. It can be seen that the STD over the slide area is greater than over the healthy area. This shows that the change of the VV backscatter on the slide part is greater than that of the non-slide areas. Although we see some high brightness pixels on the non-slide section, these are few in number compared with the slide section having a continuous cluster of high brightness pixels. Fig 3. (b, c, and d) show the horizontal, vertical, and diagonal wavelet features at level 1 or 2 for HH polarization. These figures show a good distinction between slide and non-slide pixels. As we can see, the slide

pixels have higher values in these wavelet features. This also shows that slides have higher frequency changes at different orientation.

In the next step, the SVM classifier is utilized to categorize the pixels into slide and healthy classes. SVM is a powerful supervised learning method for analyzing and recognizing the pattern and data. It tries to find a separating hyperplane in the feature space. In the SVM method, the input data are first transformed to a feature space (possibly with a higher dimension) either linearly or non-linearly based on a kernel function; then a hyperplane which separates the classes is computed by applying an optimization method [13]. In this work a polynomial kernel function with order 3 is used to map the training data into the kernel space. Also the Sequential Minimal Optimization (SMO) [14] is used to find the separating hyperplane. The parameters of the SVM such as the vector of weights and bias [13] are obtained in the training mode. Finally, in order to evaluate this methodology, the slide detection method is validated using a leave out cross validation technique..

4. VERIFICATION OF RESULTS

Fig. 4 shows a part of the levee where a slide occurred. The masks of the slide (blue) and healthy area (green) for training of the classifier are depicted in this figure. Using the method outlined in Fig. 2 and by computing the features from the data in the two masked areas, the algorithm is trained and the SVM parameters are obtained. Fig.5 shows the accuracy of classification versus different numbers of training samples. This accuracy is obtained using leave-one-out cross validation in which a single observation is used for validation and the remaining data for training.

As seen in Fig. 5, after around 80 training samples a 90% accuracy can be reached. It can be also seen that after 150 training samples, a steady accuracy of 92% is achieved. Table 1 depicts a confusion matrix based on the leave-one-out cross validation with 200 training samples (200 for both slide and healthy). The corresponding overall accuracy is around 92%. Out of the 200 pixels of slide and healthy, 23 pixels are false positive (classified healthy but they are slide pixels) and 8 are false negative (classified slide but they are healthy pixels)

Fig. 6 shows the testing of the whole area of study. As it is seen by the arrow, most of the slide pixels are detected. However, we see that detection of some non-slide pixels, which are false positives, occurs. It seems some small variation on the roughness or change on the vegetation type and height can cause the misclassification.

The results show that the foregoing method can provide good detection of landslides on an earthen levee. Although some false positive pixels are identified by the classification, these misclassified pixels are scattered and in most case they do not cluster in large segments as they do in the actual slide area.

5. SUMMARY

A method based on TerraSAR-X images and SVM was developed to detect slough slides along earthen levees. This method includes: radiometric normalization of the TerraSAR image using a high resolution DEM; feature extraction including backscatter and texture features from the levee area of interest; obtaining the SVM parameters in training mode; and classifying the pixels using the SVM in the testing mode. The roughness and corresponding textural characteristics of the soil in a slide are different from the healthy levee. Therefore, texture features along with the backscattering coefficients of the HH

and VV polarization channels are extracted. The results show a classification accuracy of 92% is reached when the number of training samples is more than 150. This remote-sensing based detection of slide pixels is expected to be significantly helpful in managing and maintaining levee systems.

Acknowledgments - This work was funded by a grant from the Department of Homeland Security Southeast Region Research Initiative (SERRI) at the Department of Energy's Oak Ridge National Laboratory. The authors would also like to acknowledge the efforts of the US Army Corps of Engineers, Engineering Research and Development Center, which collected the soil conductivity measurements used in this study. The authors also acknowledge the German Space Agency (DLR) for providing the TerraSAR-X images at scientific cost through the Office of TerraSAR-X Science Service Systems.

REFERENCES

- [1] J. V. Aanstoos, K. Hasan, C. O'Hara, S. Prasad, L. Dabbiru, M. Mahrooghy, R. A. A. Nobrega, M. Lee and B. Shrestha, "Use of Remote Sensing to Screen Earthen Levees," in *Proc. 39th IEEE Applied Imagery Pattern Recognition Workshop*, Washington, 2010.
- [2] J. V. Aanstoos, K. Hasan, M. Mahrooghy, R. A. Nobrega and P. Suarabh, "Screening of Earthen Levees Using TerraSAR-X Radar Imagery," in *4th TerraSAR-X Science Team Meeting*, Oberpfaffenhofen, 2011.

- [3] A. K. Hossain, M. A. Easson and K. Hasan, "Detection of Levee Slides Using Commercially Available Remotely Sensed Data," *Environmental and Engineering Geoscience*, vol. 12, no. 3, p. 235 – 246, 2006.
- [4] M. Mahrooghy, J. V. Aanstoos, K. Hasan and N. H. Younan, "Levee Soil Moisture Assessment Based On Backpropagation Neural Network Using Synthetic Aperture Radar," in *Proceedings of the 34th International Symposium for Remote Sensing of Environment*, Sydney, 2011.
- [5] F. T. Ulaby, R. K. Moore and A. K. Fung, *Microwave Remote Sensing: From theory to applications*, MA: Addison-Wesley, 1982.
- [6] S. Park, W. Moon and D. Kim, "Estimation of surface roughness parameter in intertidal mudflat using airborne polarimetric SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, p. 1022–1031, 2009.
- [7] Y. Oh, " Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 596- 601, 2004.
- [8] G. A. C. (DLR), "TerraSAR-X - Germany's radar eye in space," DLR, 2009. [Online]. Available: http://www.dlr.de/eo/en/desktopdefault.aspx/tabid-5725/9296_read-15979/. [Accessed 2012]
- [9] T. Fritz and M. Eineder, "TerraSAR-X Basic Product Specification Document", Bonn: DLR, 2009.
- [10] INFOTERRA, "Radiometric Calibration of TerraSAR-X Data," INFOTERRA, Friedrichshafen, 2008.

- [11] D. Small, "Flattening Gamma: Radiometric Terrain Correction for SAR Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 3081-3093, 2011.
- [12] C. S. Burrus, R. A. Gopinath and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, 1997.
- [13] J.V. N. Vapnik, *Statistical Learning Theory*, New York: Wiley-Interscience, 1998.
- [14] J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Machines*, Cambridge, MA, MIT Press, 1998.

Table Legends

Table 1. Slide/no-slide detection confusion matrix with 200 samples

Figure Legends

Fig. 1. (a) An example of a land slide on the levee; (b) a buffered HH-SAR image overlaid on a levee with a slide on Sept 15, 2010

Fig. 2. Block diagram of the slide detection algorithm

Fig. 3. Some feature values in the area of study: (a) VV-STD; (b) HH horizontal wavelet detail coefficients at level 1; (c) HH vertical wavelet detail coefficients at level 2; (d) HH diagonal wavelet detail coefficients at level 2

Fig. 4. TerraSAR-X radar data (HH polarization) overlaid with masks of slide (blue) and no-slide (green)

Fig. 5. Accuracy of slide classification using SVM versus different sampling sizes

Fig. 6. Slide classification over the levee using the slide detection algorithm

Table 1. Slide/no-slide detection confusion matrix with 200 samples

Classification/ Target	Slide	Healthy Area	Accuracy
Slide	177	23	0.88
Healthy Area	8	192	0.96
Accuracy	0.95	0.89	0.92

FIGURES

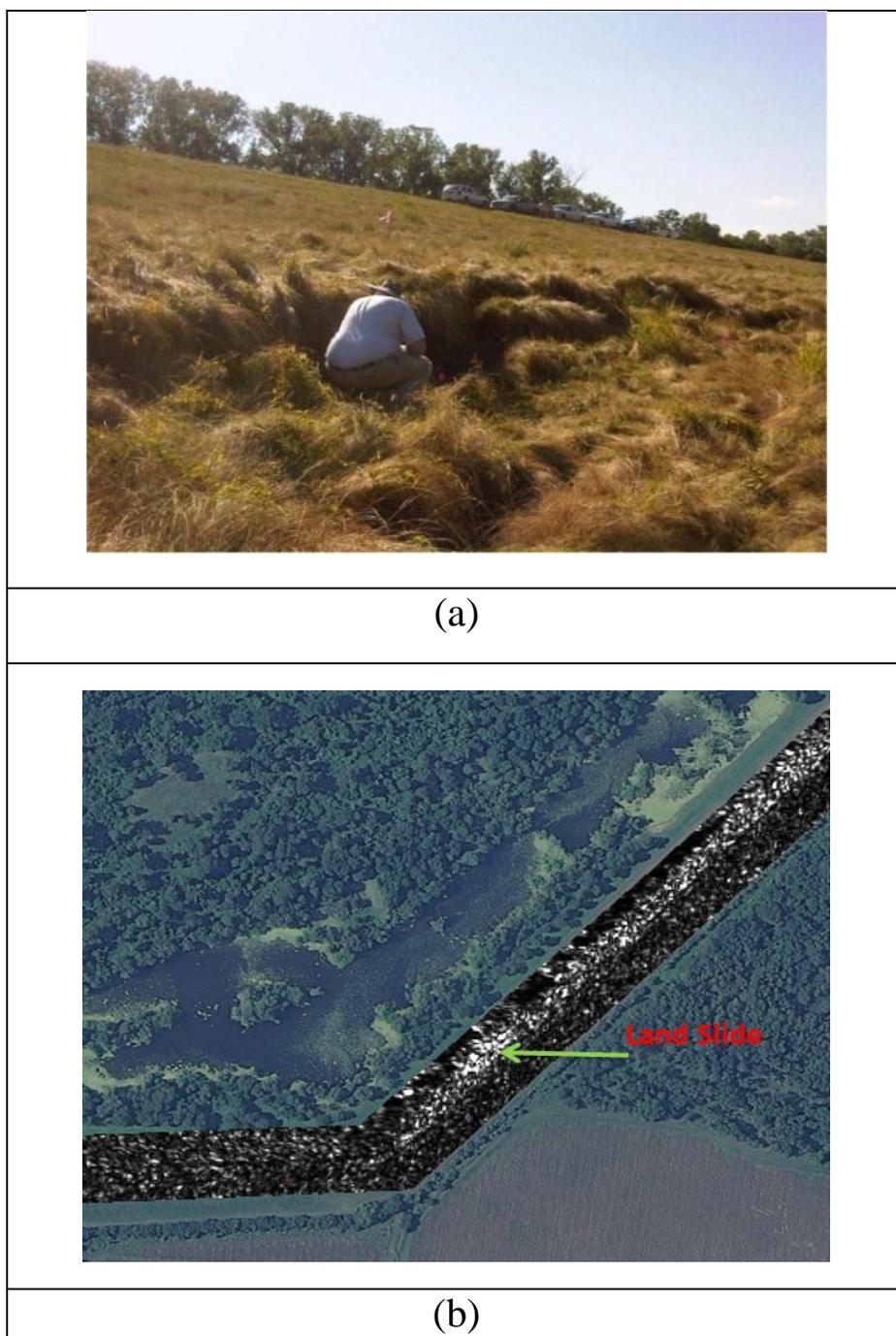


Fig. 1. (a) An example of a land slide on the levee; (b) a buffered HH-SAR image overlaid on a levee with a slide on Sept 15, 2010

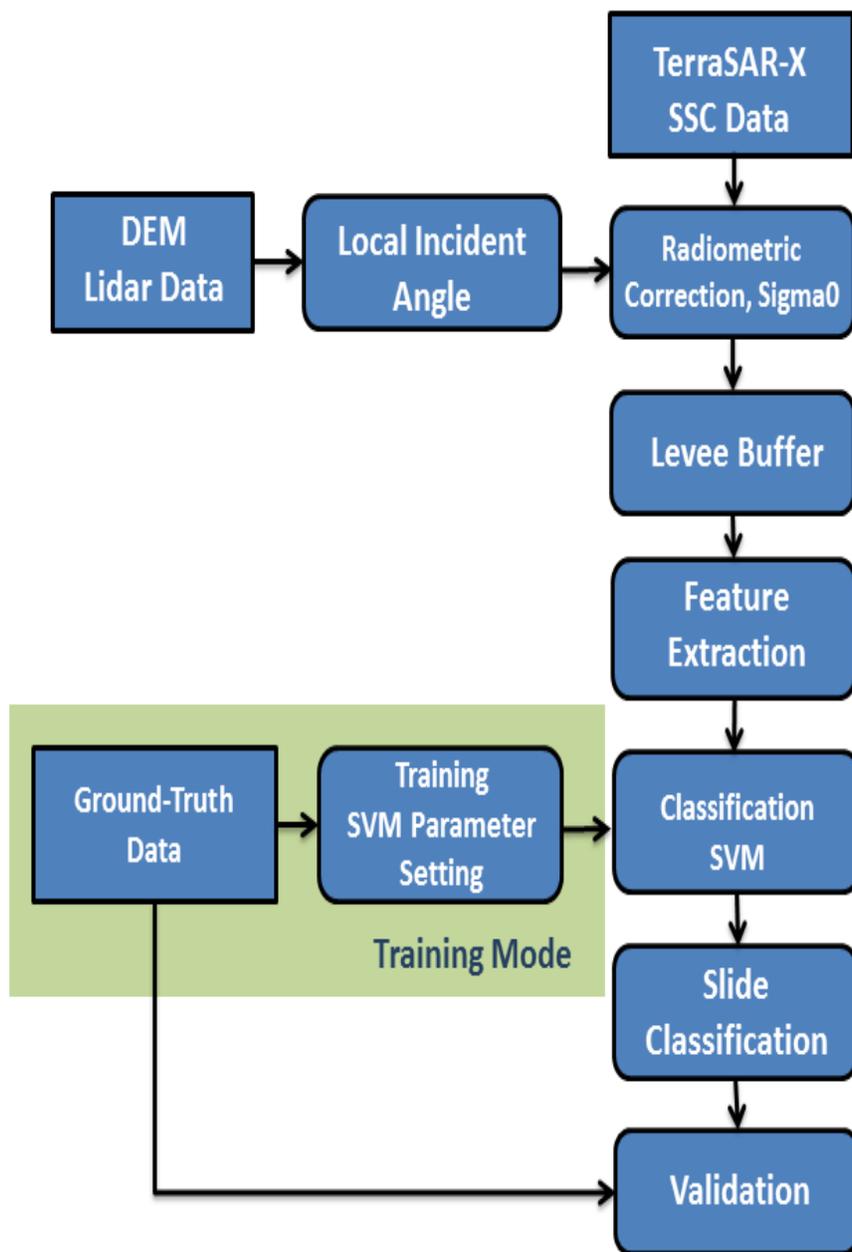


Fig. 2. Block diagram of the slide detection algorithm

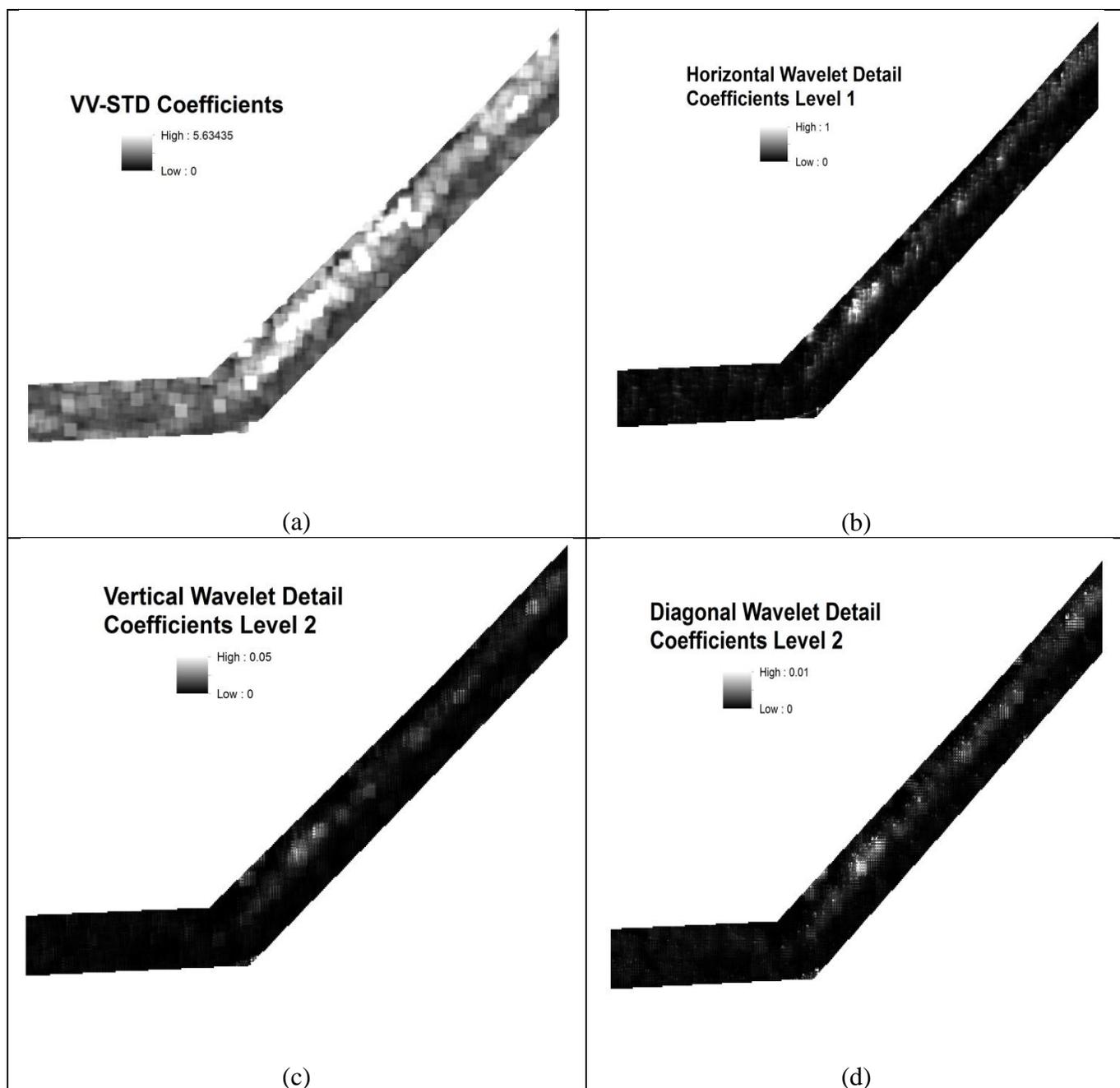


Fig. 3. Some feature values in the area of study: (a) VV-STD; (b) HH horizontal wavelet detail coefficients at level 1; (c) HH vertical wavelet detail coefficients at level 2; (d) HH diagonal wavelet detail coefficients at level 2.

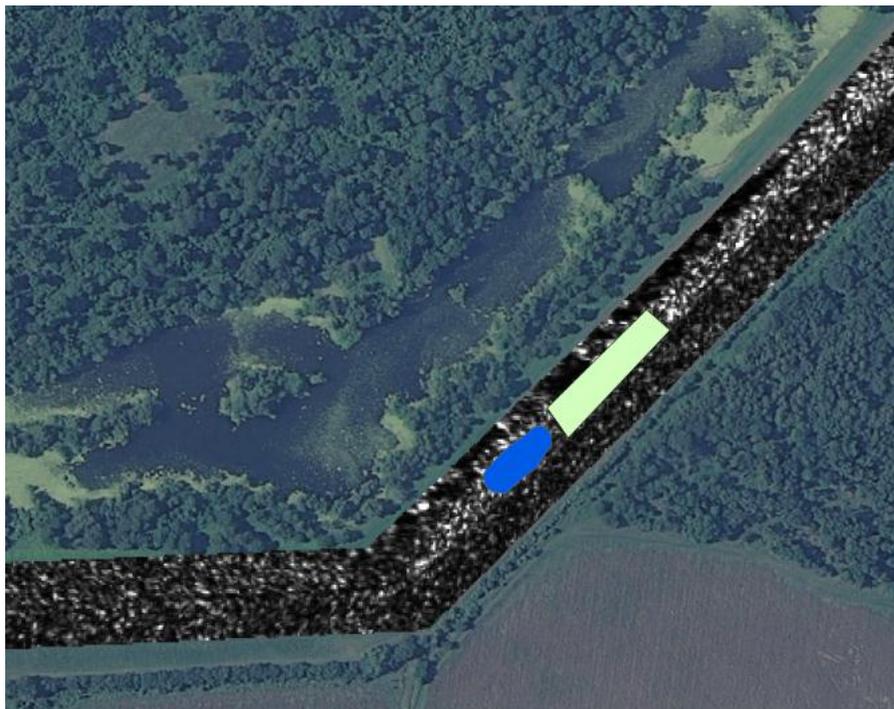


Fig. 4. TerraSAR-X radar data (HH polarization) overlaid with masks of slide (blue) and no-slide (green)

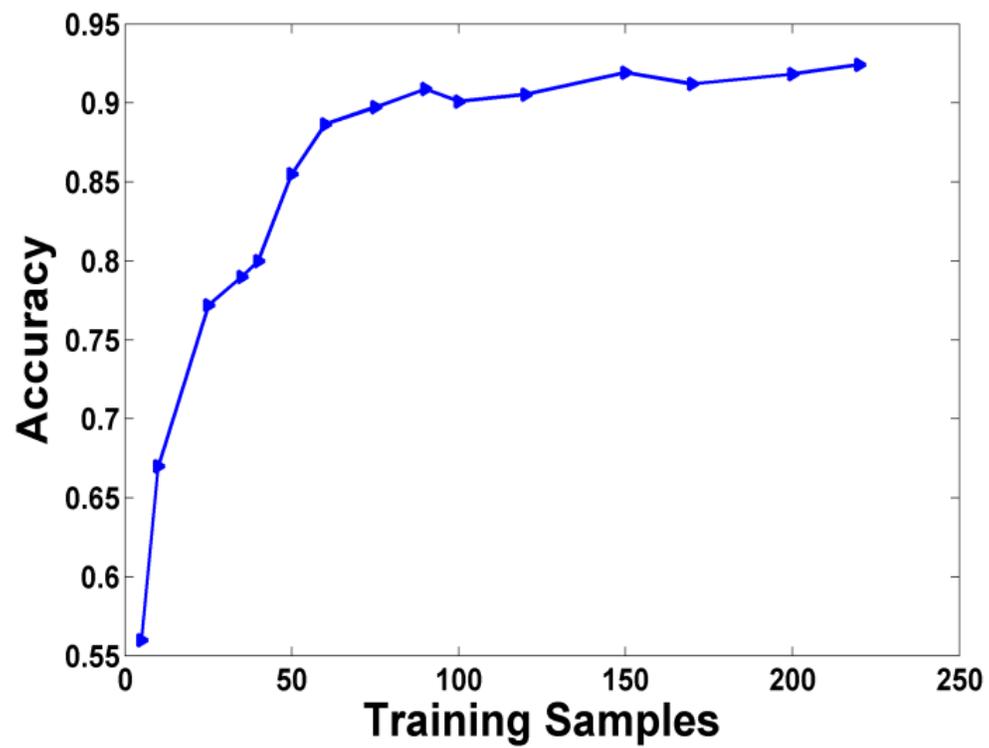


Fig. 5. Accuracy of slide classification using SVM versus different sampling sizes

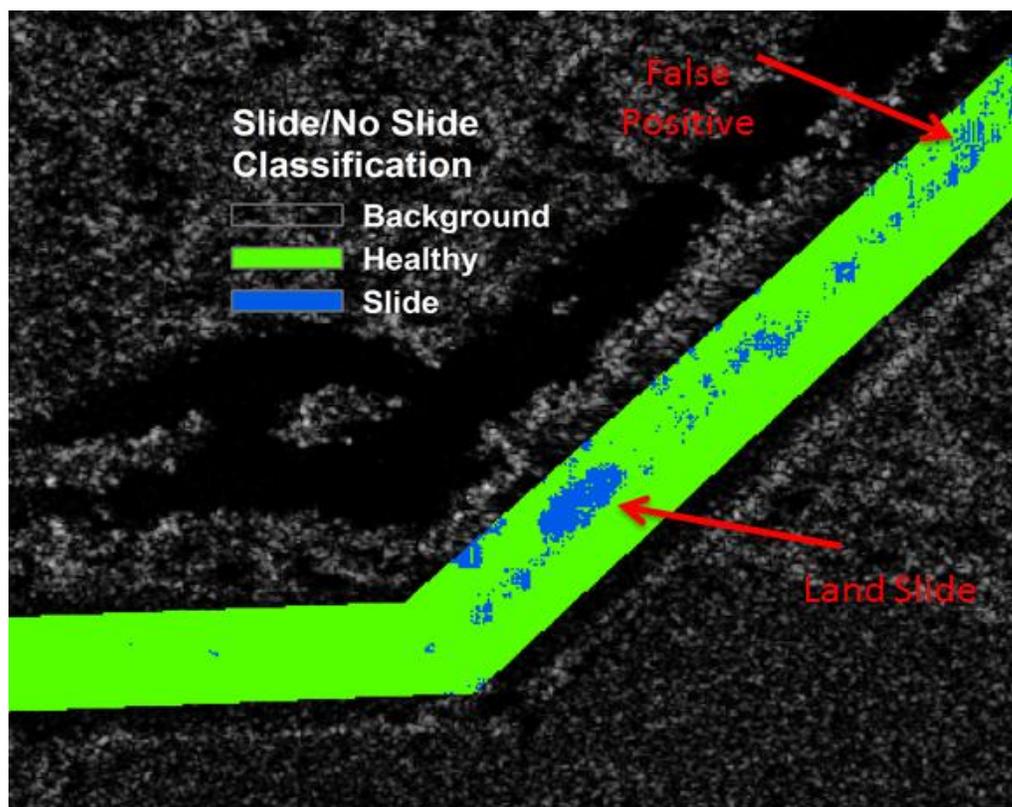


Fig. 6. Slide classification over the levee using the slide detection algorithm