12-2017

Accuracy Analysis Comparison of Supervised Classification Methods for Anomaly Detection on Levees Using SAR Imagery

Ramakalavathi Marapareddy  
University of Southern Mississippi, kala.marapareddy@usm.edu

James V. Aanstoos  
Mississippi State University, aanstoos@gri.msstate.edu

Nicolas H. Younan  
Mississippi State University, younan@ece.msstate.edu

Follow this and additional works at: https://aquila.usm.edu/fac_pubs

Part of the Computer Sciences Commons

Recommended Citation

Available at: https://aquila.usm.edu/fac_pubs/15396

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
Article

Accuracy Analysis Comparison of Supervised Classification Methods for Anomaly Detection on Levees Using SAR Imagery

Ramakalavathi Marapareddy 1,* and James V. Aanstoos 2, and Nicolas H. Younan 3

1 School of Computing, University of Southern Mississippi, Hattiesburg, MS 39406, USA
2 Geosystems Research Institute, Mississippi State University, Mississippi State, MS 39759, USA; aanstoos@gri.msstate.edu
3 Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762, USA; younan@ece.msstate.edu
* Correspondence: kala.marapareddy@usm.edu; Tel.: +1-601-266-6639

Received: 28 August 2017; Accepted: 12 October 2017; Published: 14 October 2017

Abstract: This paper analyzes the use of a synthetic aperture radar (SAR) imagery to support levee condition assessment by detecting potential slide areas in an efficient and cost-effective manner. Levees are prone to a failure in the form of internal erosion within the earthen structure and landslides (also called slough or slump slides). If not repaired, slough slides may lead to levee failures. In this paper, we compare the accuracy of the supervised classification methods minimum distance (MD) using Euclidean and Mahalanobis distance, support vector machine (SVM), and maximum likelihood (ML), using SAR technology to detect slough slides on earthen levees. In this work, the effectiveness of the algorithms was demonstrated using quad-polarimetric L-band SAR imagery from the NASA Jet Propulsion Laboratory’s (JPL’s) uninhabited aerial vehicle synthetic aperture radar (UAVSAR). The study area is a section of the lower Mississippi River valley in the Southern USA, where earthen flood control levees are maintained by the US Army Corps of Engineers.

Keywords: supervised classification; radar polarimetry; earthen levees; synthetic aperture radar (SAR)

1. Introduction

Over the entire US, there are approximately 200,000 km of levee structures of varying designs and conditions. Earthen levees (human-made earthen structures) protect large areas of populated land (hundreds of millions of Americans) and agricultural fields in the United States from flooding (as storm events increase and as river levels rise), as shown in Figure 1.

The catastrophe caused by Hurricane Katrina (August 2005) and the Midwestern United States floods (June 2008) emphasizes the importance of examining levees to improve the condition which makes levees prone to failure during floods. On-site inspection of levees is costly and time-consuming, so there is a need to develop efficient techniques based on remote sensing technologies to identify levees that are more vulnerable to failure under climate change (higher loads) and overloading of water levels (risk, flood loading) [1–3]. One type of problem that occurs along these levees which can lead to complete failure during a high-water event are slough slides [1–4] (also called landslides or slump slides). Slough slides are slope failures along a levee, which leave areas of the levee vulnerable to seepage and failure during high water events. These failures can develop over months to years, and are not quickly noticeable unless they depict some sort of visual identification [5–8]. Polarimetric synthetic aperture radar (PolSAR) data includes a variety of information which relates to the physical properties of the terrain. Hence, PolSAR provides much more information because the transmitted
signal is polarized and different polarizations of the backscatter signal are detected [9]. On the other hand, PolSAR appears to have a challenging problem in the classification field due to the complexity of extracting useful information from its multiple polarimetric channels. Nonetheless, investigations with PolSAR imagery have shown that it has great potential for classification and pattern recognition applications [10].

![Figure 1. Earthen levee at Eagle Lake, Warren County, Vicksburg, MS, USA.](image)

In the past, we have conducted work of detecting anomalies on levees using advanced unsupervised classification methods using polarimetric SAR. In that work, using the parameters entropy (H), anisotropy (A), alpha (α), and eigenvalues (λ, λ1, λ2, and λ3), we implemented several unsupervised classification algorithms for the identification of anomalies on the levee. The classification techniques applied are H/α, H/A, A/α, Wishart H/α, Wishart H/A/α, and H/α/λ classification algorithms. Additionally, we have implemented a supervised Mahalanobis distance classification algorithm for the detection of slough slides on levees using complex polarimetric synthetic aperture radar (polSAR) data [11]. The classifier output was followed by a spatial majority filter post-processing step which improved the accuracy. The dataset consisted of the multi-looked cross products, as well as individual polarization channel magnitude and phase data.

In this paper, we compare the accuracy of the supervised classification methods minimum distance (MD) using Euclidean and Mahalanobis distance, support vector machine (SVM), and maximum likelihood (ML), using SAR technology to detect slough slides on earthen levees. Our motivation for this paper is to compare the performance analysis of several supervised classification methods. SAR technology, due to its high spatial resolution and soil penetration capability, is a good choice to identify such problem areas on levees, so that they can be treated to avoid possible catastrophic failure [1,3]. The roughness and related textural characteristics of the soil in a slide affect the amount and pattern of radar backscatter. The type of vegetation that grows in a slide area differs from the surrounding levee vegetation, which can also be used in detecting slides [4]. An illustration of a slide on a levee is shown in Figure 2.

The remainder of this paper is organized as follows. Section 2 describes materials and methods, which includes data descriptions and methods used in this paper: minimum distance using Euclidean distance, minimum distance using Mahalanobis distance, support vector machine, and maximum likelihood. Section 3 describes results and discussion. Finally, Section 4 offers our conclusions.
2. Materials and Methods

The NASA Jet Propulsion Laboratory’s (JPL’s) uninhabited aerial vehicle SAR (UAVSAR) is a fully polarimetric L-band SAR which is designed to acquire airborne SAR data in a fully quad-polarimetric manner. SAR technology, due to its high spatial resolution and soil penetration capability, is a good choice to identify such problem areas on levees, so that they can be treated to avoid possible catastrophic failure [1,3]. PolSAR data contain a variety of information which relates to the physical properties of the target. In PolSAR, the transmitted signal is polarized and different polarizations of the backscatter signal are detected with different polarizations, which includes: VV (vertically transmitted and vertically received), HV (horizontally transmitted and vertically received), and HH (horizontally transmitted and horizontally received). HH and VV configurations produce like-polarized radar imagery. HV and VH configurations produce cross-polarized radar imagery. Hence, polSAR provides much more information due to transmitted signal being polarized and different polarizations of the backscatter signal are detected [12]. The L-band SAR measurements can penetrate dry soil to as much as one-meter depth. Thus, they may be valuable in detecting changes in levees that will aid in levee vulnerability classification.

The supervised classification process consists of: selection of SAR sets and data bands, image segmentation of the levee area from the water side toe to the center of the levee, training the supervised classifier with training data, testing the area of interest, and validating the results using ground truth data, as shown in Figure 3.
2.1. Data Description

The classification is performed using multi-look cross products (MLC) of the UAVSAR data acquired on 25 January 2010. The MLC datasets are real floating point, four bytes per pixel. These real powers are derived from an average of three pixels in range and 12 pixels in azimuth, i.e., the number of range looks in MLC and number of azimuth looks in MLC are 3 × 12 for each single look complex data (SLC) pixel, which correspond to HHHH, HVHV, and VVVV. Three polarization channels HHHH, HVHV, and VVVV backscatter magnitudes are used as features for the classification. The pixel spacing for the MLC data is by 7.2 m × 4.99 m for the azimuth and range directions, respectively. The pixel spacing for the SLC data is by 0.6 m × 1.66 m for the azimuth and range directions, respectively [11]. The SLC datasets (HH, HV, and VV) are oversampled in nature and are dominated by speckle noise. We chose the MLC datasets to reduce the speckle effects. We choose HHHH, HVHV, and VVVV MLC data. For the MLC data, the projected multi-look data used has a ground sample size of 5.5 m by 5.5 m. UAVSAR projects slant range images to the ground range using the backward projection method. The image segment consists of 61 × 89 × 3 pixels (columns, rows, bands), and covers a 633 m long stretch of levee in the area noticeable with a red color rectangular box on the flight line full segment radar image, overlaid on the base map, as shown in Figure 4. The full image includes the portion of boundaries of Louisiana, Mississippi, and Arkansas states. For the multi-polarized SAR imagery, it is useful to create a color composite image from the HH, HV, and VV channels that are being mapped to red, green, and blue, as shown in Figure 4, which comprises both an overview image as well as a close-up view of the segmented test area. The overview image has a swath width of 20 km and a total length of approximately 200 km. The radar is fully polarimetric with a range bandwidth of 80 MHz (resulting in better than 2 m range resolution) and flies at a nominal altitude of 13,800 m [1].
The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy). The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and visually inspect the slide areas and levee conditions. The active slide (slide 3) was present and unrepaired when identified by the Corps for slide 3, it is visible in the NAIP imagery collected in 2009 and 2010 [14], and was not repaired until after the image acquisition time on 25 January 2010. Though the date of slide appearance was not confirmed by the Corps for slide 3, it is visible in the NAIP imagery collected in 2009 and 2010 [14], and the Corps maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The availability of ground truth data for training the supervised classification processes for the present application is a challenge since the targets of interest are portions of the levee that show signs of impending failure. Once these are detected, they are quickly repaired depending on their severity [15]. The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and are well-monitored. The Corps, in association with the local levee boards, maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The availability of ground truth data for training the supervised classification processes for the present application is a challenge since the targets of interest are portions of the levee that show signs of impending failure. Once these are detected, they are quickly repaired depending on their severity [15]. The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and are well-monitored. The Corps, in association with the local levee boards, maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The availability of ground truth data for training the supervised classification processes for the present application is a challenge since the targets of interest are portions of the levee that show signs of impending failure. Once these are detected, they are quickly repaired depending on their severity [15]. The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and are well-monitored. The Corps, in association with the local levee boards, maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The river side from the center of the levee is segmented because the probabilities of occurrence of slides are greater there than on the land side, since usually the water side of the levee has steep slope compared with the land side, as we have noticed from our field trips to collect ground truth data. The slope failure is one of the main cause for the occurrence of slides and potential slides. The supervised classification methods are trained with two training classes: slide (i.e., anomalous) and non-slide (i.e., healthy) areas. We used “ground truth” reference data to train and test the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The availability of ground truth data for training the supervised classification processes for the present application is a challenge since the targets of interest are portions of the levee that show signs of impending failure. Once these are detected, they are quickly repaired depending on their severity [15]. The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and are well-monitored. The Corps, in association with the local levee boards, maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.

The availability of ground truth data for training the supervised classification processes for the present application is a challenge since the targets of interest are portions of the levee that show signs of impending failure. Once these are detected, they are quickly repaired depending on their severity [15]. The study area is one in which the levees are managed by the US Army Corps of Engineers (USACE) and are well-monitored. The Corps, in association with the local levee boards, maintains a good cumulative history of past problems and have identified particularly problematic sections of levees in the study area, as shown in Table 1. The ground truth data is provided by USACE with the approximate dates slides appeared, dates slides were repaired, and ground coordinate locations. Ground truth data was also used to train the classification algorithms. Finally, the overall slide and non-slide accuracies were computed using the confusion matrix. The study area for this work focuses on the mainline levee system of the Mississippi River along the eastern side of the river, in Mississippi, USA.
was not repaired until after the image acquisition as shown in Table 2. Hence, it was an active slide during the time of the image. Training masks were created for the slide events and labeled as either repaired or unrepaired at the time of acquisition. The training sample data from slide and non-slide (healthy) parts of the levees were obtained from the radar data using the training masks for analysis. The samples from the healthy parts of the levee near the slide events were used for training of the non-slide (healthy levee) class.

Table 1. Ground truth data from the Mississippi Levee Board.

<table>
<thead>
<tr>
<th>Slide Number</th>
<th>Length</th>
<th>Vert. Face</th>
<th>Dist. from Crown</th>
<th>Latitude North</th>
<th>Longitude West</th>
<th>Date Slide Appeared</th>
<th>Date Slide Repaired</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80'</td>
<td>2'</td>
<td>30'</td>
<td>N 32-36'-37.7&quot;</td>
<td>W 90-59'-42.3&quot;</td>
<td>Oct. 2009</td>
<td>Nov. 2009</td>
</tr>
<tr>
<td>2</td>
<td>120'</td>
<td>3'</td>
<td>15'</td>
<td>N 32-36'-32.0&quot;</td>
<td>W 90-59'-46.3&quot;</td>
<td>Aug. 2008</td>
<td>Nov. 2009</td>
</tr>
<tr>
<td>3</td>
<td>200'</td>
<td>8'</td>
<td>8'</td>
<td>N 32-36'-29.1&quot;</td>
<td>W 90-59'-48.0&quot;</td>
<td>-</td>
<td>Sept. 2010</td>
</tr>
</tbody>
</table>

Table 2. Updated slides ground truth from the Mississippi Levee Board.

<table>
<thead>
<tr>
<th>Slide No.</th>
<th>From Levee Board (8 April 2011)</th>
<th>From Visual Aerial Photo Inspection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date Slide Appeared</td>
<td>Date Slide Repaired</td>
</tr>
<tr>
<td>1</td>
<td>Oct. 2009</td>
<td>Nov. 2009</td>
</tr>
<tr>
<td>2</td>
<td>Aug. 2008</td>
<td>Nov. 2009</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>Sept. 2010</td>
</tr>
</tbody>
</table>

2.2. Methods

2.2.1. Minimum Distance Using Euclidean Distance

Minimum distance classification uses the mean vectors for each class and calculates the Euclidean distance from each unknown pixel to the mean vector for each class. The pixels are classified to the nearest class [15]. The standard deviations from mean can be no threshold, single value, or multiple values. The lower the threshold value, the more pixels that are unclassified. The maximum distance error can be no threshold, single value or multiple values. The smaller the distance threshold, the more pixels that are unclassified. The pixel of interest must be within both the threshold for distance to the mean and the threshold for the standard deviation for a class. The classification is performed with zero standard deviation from the mean, and zero maximum distance error, to classify all pixels. Minimum distance classification calculates the Euclidean distance for each pixel in the image to each class [15]:

\[ D_i(x) = \sqrt{(x - m_i)^T(x - m_i)} \]  

where,

\( D = \) Euclidean distance

\( i = \) the \( i \)th class

\( x = n\)-dimensional data (where \( n \) is the number of features)

\( m_i = \) mean vector of a class

2.2.2. Minimum Distance Using Mahalanobis Distance

The Mahalanobis distance is a direction-sensitive distance classifier that uses statistics for each class in a manner like the maximum likelihood classifier. However, it assumes all class covariances are equal. Therefore it is a faster method and weighing factors are not required [15]. All pixels are classified to the closest training data. It is largely based on a normal distribution of the data in each band which is used as input to classification. The maximum distance error can be a zero threshold for all the classes, or single value for all the classes, or different values for individual classes. The distance
threshold is the distance within which a class must fall from the center or mean of the distribution for a class. We used a zero threshold for all the classes and zero maximum distance error, to classify all pixels. The Mahalanobis distance classification calculates the distance for each pixel in the image to each class using the following equation [15]:

\[
D_i(x) = \sqrt{(x - m_i)^T \Sigma_i^{-1} (x - m_i)}
\]

(2)

where:

\(D = \) Mahalanobis distance
\(i = \) the \(i\)th class
\(x = \) \(n\)-dimensional data (where \(n\) is the number of features)
\(\Sigma_i^{-1} = \) the inverse of the covariance matrix of a class
\(m_i = \) mean vector of a class

2.2.3. Support Vector Machine

Support vector machine (SVM) is a supervised classification method derived from statistical learning theory that often yields good classification results from complex and noisy data [15–18]. SVM separates the classes with a decision surface that maximizes the margin between the classes [19]. The benefit of SVM is that it works well with small training datasets. A kernel type must be selected for use in the SVM classifier. These kernels show a critical role in SVM classification. The kernel types are linear, polynomial, radial basis function, and sigmoid. If the kernel type is polynomial, we need to choose the degree of the kernel polynomial. If the kernel type is polynomial or sigmoid, we need to specify the bias in the kernel function. If the kernel type is polynomial, radial basis function, or sigmoid, we must also specify a gamma value which defines how far the influence of a single training sample reaches [19]. The kernel type used here is a radial basis function. Gamma in the kernel function used is 0.333 (ideally the gamma value is the inverse of bands used in the input image, the number of bands used in this data, were 3. So, 1/3 = 0.333). The penalty parameter, which controls the trade-off between allowing training errors and forcing rigid margins, used is 100. The pyramid levels, which set the number of hierarchical processing levels to apply during the SVM training and classification process, used is 0.00, to process the images at full resolution. We use the classification probability threshold to set the probability that is required for the SVM classifier to classify a pixel. The classification probability threshold used is zero, to classify all pixels.

2.2.4. Maximum Likelihood

Maximum likelihood 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 the class that has the highest probability (i.e., the maximum likelihood) [15]. The threshold is a probability minimum for inclusion in a class. We used a zero threshold for all the classes to classify all pixels. Maximum likelihood classification calculates the following discriminant functions for each pixel in the image:

\[
g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)^T \Sigma_i^{-1} (x - m_i)
\]

(3)

where:

\(i = \) the \(i\)th class
\(x = \) \(n\)-dimensional data (where \(n\) is the number of features)
\(p(\omega_i) = \) probability that a class occurs in the image and is assumed the same for all classes
\(|\Sigma_i| = \) determinant of the covariance matrix of the data in a class
3. Results and Discussion

The minimum distance (MD) using Euclidean and Mahalanobis distance, support vector machine (SVM), and maximum likelihood (ML) supervised classification processes were run separately, with the polSAR multi-looked cross product data. The test study area SAR image and corresponding geographically-located optical image, overlaid with the levee center line, are shown in Figure 5. Using the ground truth data, image masks were created bounding the active slide area and a subset of the non-slide area within segmented image (i.e., test area), as shown in Figure 6. A sample of the pixels in of each of these two classes was then used to train all four supervised classifiers. The accuracy of the resulting classification was tested using the remaining reference data pixels for testing, and the conventional statistics of user, producer, and overall accuracy were computed for each method (in the form of percentages as well as pixel) using the confusion matrix.

\[ \Sigma_i^{-1} = \text{the inverse of the covariance matrix of a class} \]

**Figure 5.** (a) Test area SAR image; and (b) optical image; overlaid with the levee center line.

**Figure 6.** (a) Segmented test area SAR image; and (b) classification training classes; overlaid with the levee center line.
The classification maps resulting from applying the MD using Euclidean and Mahalanobis distance, SVM, and ML classification to test areas using co-polarized features (HHHH, HVHV, VVVV), are shown in Figures 7–9. The training masks are shown in Figure 6b for both slide and non-slide classes. The accuracy assessment results for all four classifiers are tabulated in Table 3. The summary of accuracy assessment results which include overall accuracy, slide accuracy, and non-slide accuracy are listed in the Table 4. The training masks, shown in Figure 6b, cover 78 and 84 pixels for the slide and non-slide area, respectively. Total pixels of the testing area (as shown in Figure 6a) were 929. Of these, 17% (162 pixels) were used for training the classifier and the remainder used for testing its accuracy. All 4 supervised classification results show good detection of the slide areas, but numerous false positive detections as well. The test area included one active slide and two slides (numbered 1 and 2, as shown in Figure 9) which had been repaired by the time of image acquisition. Many of the false positive pixels fall in this repaired area. Since these slide areas were repaired only two months prior to the time of image acquisition, they still have characteristics more like the active slide than the “healthy” areas in terms of surface roughness and differences in the grass cover. These characteristics likely influenced the classification. Out of these two repaired slides, the slide 2 area has more false positives, since it appeared in August 2008, and was repaired in November 2009. To the contrary, the slide 1 area has fewer false positives since it appeared in October 2009, and was repaired in November 2009. Hence, the slide 2 areas might contain more surface roughness and differences in the grass cover. The classification accuracies obtained for all four methods are quite high. The ML results are the best, followed by SVM, and MD using Mahalanobis and Euclidean distance. All four methods achieved near-perfect identification of non-slide pixels, but differ in the detection of slide pixels (thus, in their false positive results, identifying some non-slide pixels as slides). In addition to the active slide areas, other anomalous areas are also detected to some degree and classified into the slide class. Some of these are previous slide areas that had been repaired just two months prior to the time of image acquisition and still appear similar enough to the active slide to be detected by the classification technique.

![Figure 7](image-url)  
*Figure 7. Minimum distance classification, using: (a) Euclidean distance; and (b) Mahalanobis distance.*
Figure 8. (a) Support vector machine classification; and (b) maximum likelihood classification.

Figure 9. (a) Maximum likelihood classification; and (b) optical image; overlaid with the levee center line, slide, and non-slide areas.
Table 3. Accuracy analysis of the MD using Euclidean and Mahalanobis distance, SVM, and ML classification, for the slide and non-slide areas.

<table>
<thead>
<tr>
<th>Classification Method Vs. Accuracy (%)</th>
<th>Overall Accuracy = (132/162) 81%</th>
<th>Ground Truth (Pixels)</th>
<th>Ground Truth (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>slide3</td>
</tr>
<tr>
<td>MD u/Euclidean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>slide3</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>non-slide</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>78</td>
</tr>
<tr>
<td>MD u/Mahalanobis</td>
<td></td>
<td>Total</td>
<td>78</td>
</tr>
<tr>
<td>Overall Accuracy = (146/162) 90%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVM</td>
<td></td>
<td>Overall Accuracy = (158/162) 97%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>slide3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>non-slide</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>78</td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td>Total</td>
<td>78</td>
</tr>
<tr>
<td>Overall Accuracy = (159/162) 98%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Accu. ≥ Accuracy; Prod. ≥ Producer; u/Euclidean ≥ using Euclidean; u/Mahalanobis ≥ using Mahalanobis.

Table 4. Accuracy analysis of the MD using Euclidean and Mahalanobis distance, SVM, and ML classification, for the slide and non-slide areas.

<table>
<thead>
<tr>
<th>Classification Method Vs. Accuracy (%)</th>
<th>Overall Accuracy</th>
<th>Slide Accuracy</th>
<th>Non-slide Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD u/Euclidean</td>
<td>81</td>
<td>61</td>
<td>100</td>
</tr>
<tr>
<td>u/Mahalanobis</td>
<td>90</td>
<td>79</td>
<td>100</td>
</tr>
<tr>
<td>SVM</td>
<td>97</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>ML</td>
<td>98</td>
<td>97</td>
<td>98</td>
</tr>
</tbody>
</table>

4. Conclusions

Supervised classification methods based on the minimum distance (MD) using Euclidean and Mahalanobis distance, support vector machine (SVM), and maximum likelihood (ML) classification for earthen levee slide detection using UAVISAR imagery were presented. The effectiveness of the algorithms is demonstrated using fully quad-polarimetric L-band SAR imagery from the NASA JPL’s UAVISAR. Three polarization channels HHHH, HVHV, and VVVV back-scatter magnitudes are used as features for all four supervised classification methods. The study area is a section of the lower Mississippi River valley in the Southern USA. The classification results obtained for all four procedures indicate that the use of polarimetric SAR can effectively detect slide areas on earthen levees. The overall performance of the ML classification was the best, but all had quite good accuracy. The SVM method is somewhat more complicated to implement than the other two methods, involving more parameters and options which must be specified. Although the test study area is small, including only one active slide area and two anomalous areas for testing, the methodologies presented in this paper show promising results. Planned future work includes the use of larger test areas consisting of more active slides, seasonal images acquired by the SAR, and different geometrical orientations of the levee.
Acknowledgments: This work was supported by the National Science Foundation grant number: OISE-1243539, and by the NASA Applied Sciences Division under grant number: NNX09AV25G. The authors would like to thank the US Army Corps of Engineers, Engineer Research and Development Center and Vicksburg Levee District for providing ground truth data and expertise; NASA Jet Propulsion Laboratory for providing the UAVSAR image.

Author Contributions: Ramakalavathi Marapareddy implemented the classification methods. James V. Aanstoos supervised the work, provided imagery and data, and was the principal investigator for the project. Nicolas H. Younan supervised and provided guidance. Ramakalavathi Marapareddy, James V. Aanstoos, and Nicolas H. Younan analyzed the results and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Dunbar, J. USACE’s lower Mississippi valley engineering geology and geomorphology mapping program for levees. In Presentation at Vicksburg; USACE: Vicksburg, MS, USA, 16 April 2009.