Automatic CRP Mapping and Rectification using Nonparametric Machine Learning Approaches

Xiaomu Song, Student Member, IEEE, Guoliang Fan, Member, IEEE, and Mahesh Rao

Abstract—This paper studies an uneven 2-class classification problem of satellite imagery, i.e., the mapping of United States Department of Agriculture (USDA)‘s Conservation Reserve Program (CRP) tracts. CRP is a nationwide program that encourages farmers to plant long-term resource conserving covers to improve soil, water and wildlife resources. With the recent program development, it is imperative to obtain accurate CRP maps or rectify existing CRP reference data for management and evaluation purposes. Since the CRP mapping is a complex classification problem where both CRP and non-CRP areas are composed of various cover types, we introduce multisource data classification in this work. Two nonparametric machine learning approaches, i.e., decision tree classifier (DTC) and support vector machine (SVM) are implemented. Specifically, considering the importance of CRP mapping sensitivity, a new DTC pruning method is proposed to increase the recall rate. We also study two relaxation approaches to increase the recall rate of SVM. Moreover, a localized and parallel framework is suggested in order to efficiently deal with the large scale CRP mapping and rectification problem. Simulation results validate the applicability of suggested framework and proposed techniques.

Index Terms—Multisource data classification, decision tree, support vector machine, conservation reserve program.

I. INTRODUCTION

The classification of remotely sensed imagery into different areas for Land Use Land Cover (LULC) analysis has been an important topic over the last decades. Conventional parametric statistical model based methods show their efficiency in such problems [14], [29]. In recent years, a variety of work use multisource geospatial data to facilitate the classification of multispectral imagery [25], [23], [42], [8]. Correspondingly, people found that it may not be appropriate to model multisource data by traditional multivariate statistical models [25], [23], [5], [30], [4]. Therefore, nonparametric methods should be considered. In this work, we study nonparametric machine learning approach for mapping and rectifying United States Department of Agriculture (USDA)’s Conservation Reserve Reserve Program (CRP) tracts based on satellite imagery, which is a special and complex problem of LULC analysis. CRP is a program that encourages farmers to plant long-term resource conserving covers to improve soil, water and wildlife resources [1]. Very little work has been done for CRP mapping so far, and recent work in [15] requires considerable human interpretation and intervention.

Compared with the traditional LULC applications, CRP mapping has several major characteristics that make it a complicated problem. (1) CRP mapping is a 2-class classification problem (CRP and non-CRP) of complex rural area where each class is a mixture of many different land cover types, resulting in highly overlapped clusters in the spectral spaces of satellite imagery. Therefore, representative feature sets and powerful data classifiers are necessary. (2) Existing CRP reference data is not very accurate, and we should find a specific way to select reliable training samples and to evaluate the mapping performance from present reference data. Moreover, based on the mapping results, we can also rectify some errors in the reference data. (3) CRP mapping is an uneven classification task where CRP tracts usually count less than 10% in a study area. Additionally, due to requirements of compliance monitoring, methods that favors high recall rates should be considered. (4) Since CRP is a nationwide program, any CRP mapping tools developed should be computationally efficient with minimum human involvement.

Basically, CRP is a man-made program that possesses strong correlation with geographic information system (GIS) data, such as slope, elevation and distance-to-waterbody, therefore the importance of the multisource GIS data is particularly prominent in the CRP mapping problem. An insightful understanding of the use of multisource data is that if a proper number of training samples are available, the cluster separability of different classes can be increased in a higher dimensional feature space constructed from multisource data. The work in [8] shows the advantage of neural network on the classification of complex rural areas. In this work, we study decision tree classifier (DTC) and support vector machine (SVM) [44] for CRP mapping and rectification. The principle of DTC is to break up a complex classification problem into a union of several simpler classification issues. SVM constructs a linear classification hyperplane that maximize the margin between two different training patterns in the original feature space or a high dimensional feature space generated by kernel methods [44], [6], [12], [9]. In our recent work [43], DTC and SVM were used to implement CRP mapping based on multisource geospatial data, where the CRP reference data was used as ground truth for performance evaluation. However, although there is usually no significant mis-location of CRP tracts in the reference data, some locality error around CRP boundaries could deteriorate the purity of training sample, and invalidate the performance evaluation. In this work, we use a specific approach to refine the classification training and to evaluate the mapping performance. As the by-product, we can also rectify the present CRP reference data.
Since the CRP mapping is an uneven 2-class classification problem, besides the overall classification accuracy, precision (user’s accuracy) and recall (producer’s accuracy) are used to evaluate the overall CRP mapping performance. Usually, both pruned DTC and SVM favor high precision, leading to low recall if the classification accuracy cannot be increased significantly. However, failing to detect existing CRP tracts is more undesirable than false CRP tracts when we deal with compliance monitoring, thus high recall outweighs high precision in practical CRP mapping. Therefore, we propose a new DTC pruning method to increase recall. We also study two relaxation approaches for SVM to improve recall specifically. Moreover, we propose a localized and parallel classification framework to implement CRP mapping for large areas efficiently and effectively.

More details of above issues will be discussed in the following sections. In Section II, USDA’s CRP program and the study area will be introduced. DTC and SVM are briefly described in Section III. In Section IV, the localized framework is proposed based on the multisource geospatial data. Section V studies how to improve the sensitivity of DTC and SVM for CRP mapping. Section VI shows and discusses simulation results. Conclusions are drawn in Section VII.

II. THE CRP PROGRAM AND STUDY AREA

This work is originally motivated by the need for mapping USDA’s CRP tracts from remotely sensed data. CRP as a provision of the 1985 Farm Bill seeks to convert highly erodible lands with active crop production to permanent vegetative cover [11]. It is a voluntary program that uses financial incentives to encourage farmers to enroll in contracts of 10-15 years in duration to remove lands from agricultural production. Enrolled lands must be highly erodible, contribute to a serious water quality problem, or provide substantial environmental benefits if devoted to certain specific conservation uses. USDA’s Farm Service Agency (FSA), in-charge of administering the CRP signups and enrollments, evaluates the fields submitted by the producers based on the Environmental Benefit Index (EBI) score accumulated by each farm applicant (FSA 2003). This process implicitly associates CRP enrollments with multisource GIS information. Depending on the overall applicants, a cutoff EBI is identified, above which the farms get selected for the long-term retirement with rental benefits. Starting in 1998, with the initial CRP contracts beginning to expire and a nearly 1.6 billion new enrollment in 2003, it is imperative for FSA to evaluate and manage this program based on accurate and detailed digital CRP maps, which are usually not available or need to be updated in terms of mapping accuracy.

Currently, there is no standardized approach to keeping track of existing CRP tracts. FSA relies on aerial photography to manually delineate CRP tracts on the county level basis. These aerial photographs are at the section level, and provide little information about the CRP from a landscape perspective. Furthermore, when Natural Resources Conservation Service (NRCS) drafts the CRP reference data, the possible mislocality and spatial misalignment of CRP tracts deteriorate the reliability and usability of these reference data [15]. Therefore, the goal of this work is to develop an automatic tool for accurate CRP mapping and rectification based on the existing CRP reference data provided by NRCS.

The study area of this work is located in Texas County, Oklahoma as shown in Fig. 1 (a). This area (552 × 523 pixels) is about 260km², where the accurate CRP mapping and rectification framework is developed and tested. Fig. 1 (b) illustrates the imperfect CRP reference data of this area. Texas County is one of the most intensively farmed counties in Oklahoma. Because of the underlying water-rich Ogallala Aquifer, irrigated farming is extensively practiced in the area for corn, sorghum, cotton, and soybeans cultivation. Due to the large scale of agriculture for many years, Texas County also ranks first in the state for CRP enrollments. Therefore, it is a typical region for the study of CRP mapping and rectification.

III. MACHINE LEARNING APPROACHES

As mentioned before, nonparametric machine learning approaches are used CRP mapping. Machine learning is the ability of a machine to recognize patterns that have occurred repeatedly and to improve its performance based on past experiences. It is a typical machine learning problem that aims at acquiring general concepts for two different land covers, i.e., CRP and non-CRP, from given training samples. The target function of CRP mapping is defined as:

\[ Y = f(X), Y \in \{0, 1\}, \]  

where \( X \) is the multisource geospatial data, \( f(\cdot) \) is the target concept to be learned, and \( Y \) is an indicator where 1 can be defined as CRP and 0 as non-CRP. Both DTC and SVM are inductive inference methods, and the learning goal is to determine a hypothesis of target concept to best fit the training data. DTC has been shown advantages in real remote sensing (RS) applications for more than ten years [23], [3], [41], [17], [13], [21]. However, considering the overfitting problem met by DTC with poor generalization performance, SVM is suggested as an alternative to DTC. Recent research on SVM in RS applications have shown impressive classification results [20], [19], [18], [22], [32], [35], [7]. In this work, both DTC and SVM are implemented for CRP mapping as a semi-supervised classification issue involving multisource
geospatial data. Particularly, the generalization performance of two machine learning approaches is carefully studied to produce accurate CRP maps with high recall rates.

A. Decision Tree Classifier (DTC)

DTC is a tree-structured classifier built from a training data set, representing rules underlying training data with hierarchical and sequential structures that recursively partition the data. In this work, the C4.5 DTC is applied to the CRP mapping [38]. It is constructed based on the information gain ratio criterion, which measures the increase in class purity. Assuming a set of samples \( S \) that contains \( k \) classes with probability \( p_1, \ldots, p_k \), if \( S \) is partitioned into \( n \) classes based on a test, the information gain ratio is defined as:

\[
\text{Gain-ratio}(S) = \frac{\text{Gain}(S)}{\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log\left(\frac{|S_i|}{|S|}\right)},
\]

where

\[
\text{Gain}(S) = \text{Info}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \text{Info}(S_i),
\]

and

\[
\text{Info}(S) = -\sum_{i=1}^{k} p_i \log(p_i).
\]

In equation (2), \(|S_i|\) is the number of samples in subset \( i \) and \(|S|\) is the number of samples in the set \( S \). Gain(S) is the gained information of the target function that is obtained from the test with selected features, and Gain-ratio(S) is a normalized information gain so that the bias of trivial partition could be avoided [37]. Beginning from the root node, the C4.5 performs a top-down greedy search through the complete hypothesis space until the stop criterion is met. In the simulation, the tree stops growing if there are less than five samples in a node.

B. Support Vector Machine (SVM)

SVMs are newly developed learning methods [44]. Given a set of training samples from two classes: \( \{(x_1, y_1), \ldots, (x_l, y_l)\} \), \( x \in \mathbb{R}^n \), \( y \in \{1, -1\} \), the goal of SVM learning is to determine an classification hyperplane induced from the training samples that maximally separates classes, or equivalently, to minimize \( \frac{||w||^2}{2} \), subject to

\[
y_i((w \cdot x_i) + b) - 1 \geq 0, \quad y_i \in \{1, -1\}, \quad \forall i.
\]

where \( w \) and \( b \) are parameters of the hyperplane. If the training data are linearly nonseparable, the hyperplane can be obtained by minimizing:

\[
C \sum_{i=1}^{l} \xi_i + \frac{1}{2} ||w||^2,
\]

subject to :

\[
y_i((w \cdot x_i) + b) \geq 1 - \xi_i,
\]

where \( \xi_i \geq 0, \quad i = 1, \ldots, l \) is called slack variables, and \( C \) indicates the tradeoff between the complexity of classification hyperplane and the ratio of nonseparable data samples.

In SVM learning, kernel methods are often used to map the data vectors in the input space into a higher dimension feature space, then the construction of a linear classification hyperplane in this high dimension feature space is equivalent to a nonlinear decision hyperplane in the input lower dimension space [6], [9]. There are several often used kernel functions, such as the radial basis function (RBF):

\[
K(x_i, x_j) = \exp(-||x_i - x_j||^2 / 2\sigma^2),
\]

where \( \sigma \) is related to the function width. In this work, we use a nonlinear SVM with a RBF kernel, and a SVM software SVMlight [27] is used to perform training and classification of multisource geospatial database.

C. Performance Measurements

The generalization performance is one of the most important issues of machine learning approaches because it shows how well the learned hypothesis approximates the true target concept. Regarding the CRP mapping performance, three measurements are used in this work: classification accuracy is defined as the percentage of pixels that are correctly classified in terms of CRP and non-CRP. Precision indicates the percentage of detected CRP pixels that are true ones. Recall is the percentage of true CRP pixels that can be detected. A recent research reveals that any classification system that performs better than a random decision exhibits a tradeoff between precision and recall if classification accuracy is a constant [2]. This implicates that if we cannot further increase classification accuracy, we could only improve precision by sacrificing recall, and vice versa. A further increase of precision and recall could not happen simultaneously unless classification accuracy could be increased, which is difficult and costly. Therefore, searching for a proper tradeoff is more realistic in the case of CRP mapping. According to [2], a tradeoff between precision and recall is formulated as:

\[
\lambda P_c + (\lambda + P_a - 1)P_b = 2\lambda P_a P_c,
\]

where \( P_a \) is classification accuracy, \( P_b \) is precision, \( P_c \) represents recall, and \( \lambda \) indicates the probability that a randomly selected sample belongs to the class of interest. Since CRP mapping is an uneven classification problem, where CRP tracts might cover less than 10% of a whole study area, i.e., \( \lambda \approx 0.1 \) in most cases, we hope that the trained classifiers can achieve high \( P_c \) for testing data. We will discuss how to increase \( P_c \) for DTC and SVM in Section V.

IV. CLASSIFICATION FRAMEWORK

A. Multisource Geospatial Database

The multisource geospatial database is composed of the Landsat TM satellite imagery, derived features, texture information and GIS data, which are all in the raster format. Combined with LULC GAP data and CRP reference data, there are a total of 52 layers as shown in Fig. 2. During CRP mapping, layers A and D are original inputs, and layers B and C are automatically generated by the system during run time.
Layer A is the multi-temporal Landsat TM imagery with a resolution of $30m \times 30m$, including imageries acquired in February and June, 2000. Bands 2, 3, 4, 5, and 7 for the two seasons are used in the study, resulting in the first 10 layers of the database from top to bottom.

Layer B contains derived features from satellite imagery, including Normalized Difference Vegetation Index (NDVI), band ratio and band difference between different seasons. The NDVI for Landsat TM is computed as:

$$NDVI = \frac{Band_4 - Band_3}{Band_4 + Band_3} \quad (8)$$

where $Band_3$ and $Band_4$ are spectral values in bands 3 and 4, respectively. The NDVI is calculated from the imagery in each season and the largest one is chosen as the final value. It can be used to discriminate different vegetation cover types. We also apply (8) to other image bands to generate more layers of vegetation indices. Band ratio is defined as:

$$BR_{i,j} = \frac{Band_i}{Band_j} \quad (9)$$

where $i$ and $j$ is the index of Landsat TM image bands. The band ratio can be used for reducing environmental interference on remotely sensed imagery [39], [26], [31], such as the effect of solar elevation, and providing information for discrimination of soils and vegetation cover types, which is not available in any single band image. Band difference is the difference in the same band from two different seasons:

$$BD_{m,n}^i = Band_m^i - Band_n^i \quad (10)$$

where $m$ and $n$ indicate different seasons of band $i$. Band difference could provide information on discriminating seasonal changing areas, e.g., croplands and grasslands, and unchanging areas, e.g., urban, bare soil, etc. Totally, there are 17 layers in the database that are composed of NDVI, band ratio and band difference data.

Layer C consists of 20 layers of texture information, including local mean and local variance of each band in different seasons. The local mean is the average spectral value within a $3 \times 3$ window, and the local variance is computed based on the local mean in the same window. The texture layers are followed by GIS data of Layer D, including elevation that ranges from 881 to 986 feet, slope that is from 0 to 30, and distance-to-waterbody with the extent from 0 to 3230 feet. The LULC GAP data could be used for more robust image analysis with respect to different cover types. The bottom layer is the reference data for training and/or evaluation purposes.

### B. Feature Extraction

The multisource geospatial database is directly applied to DTC to generate a set of rules that can be easily understood by researchers. Before implementing SVM, since different layers of the database have different units, it is necessary to normalize them so that one feature cannot dominate others. Each layer of the database is normalized via subtracting the mean and being divided by the standard deviation of the layer. Moreover, the Landsat TM spectral channels, the derived features and texture information from these channels contain redundant information, and it is necessary to reduce the feature dimension via feature selection or extraction. In this work, we use conventional principle component analysis (PCA) to extract feature subsets that preserve 98% variation, resulting in 5-8 feature layers out of 47 layers. Including 3 GIS layers, there are totally about 10 layers for SVM-based GIS layers.

### C. Localized Data Classification

When selecting training samples, a straightforward way is to select samples that cover all the study area. Nevertheless, it is worth pointing out that there are more than 30 different CRP species in Texas County (14 CRP species in the study area), and more cover types in non-CRP regions, such as crop, urban, and pasture, etc. The 3-D spectral distributions of CRP and non-CRP areas in the study area are illustrated in Fig. 3 (a). It can be seen that the spectral features from CRP and non-CRP regions highly overlap in bands 3, 4, and 7. One possible approach to reduce the overlapping is to split CRP and non-CRP areas into many single cover types, and select training samples from them individually. However, this will introduce other problems: first, some CRP species have small coverage, and enough training samples may not be available. Second, detailed cover types are not available, and some CRP species are still mixed grasses. LULC GAP data cannot provide such information. Thirdly, some CRP species also overlap in the feature space as shown in Fig. 3 (b). Fourthly, when the study area is at the county level, the total number of training samples might be very large in order to guarantee that every cover type is included. This will increase the computational load significantly. In this work, we use a localized process to achieve automatic and accurate CRP mapping/rectification efficiently. This process splits the study area into smaller blocks, and the nonparametric data classifiers are trained by given training samples in each block, and learned knowledge is applied to classify all test samples in this block. The output of all blocks can be combined to rebuild the whole CRP map. The classification framework is shown in Fig. 4.

Three major reasons support the block-wise localized operation for CRP mapping. (1) Less cover types exist in each block and the overlap between CRP and non-CRP in the feature space could be reduced. (2) Enough reliable training samples (above 50% of the true CRP) are usually available from the existing reference data for each block. (3) The localized process leads to a highly parallel classification structure, reducing computation time dramatically. For example, when the sampling rate is 20%, which means 20% of CRP and non-CRP areas are used as the training sample and the remained
areas are selected as the test sample, it takes more than half an 
hour to training a SVM and perform classification (Pentium 
IV 2.2GHz CPU, 1G bytes memory) on the whole study 
area. If the study area is split into 25 blocks of size around 
100 × 100, the time for SVM training and classification in 
each block is about 10 seconds, and 30 seconds for DTC. If 
we have a parallel computing architecture, the whole area can 
be processed efficiently.

In the study area, we found that there is weak dependence 
cross all blocks and the learned data classifier in a block may 
not be applicable to others. This is not unexpected, because 
different regions usually contain different CRP grass species 
and non-CRP sites as well as distinct GIS attributes. We 
manifest this fact by training a SVM in a randomly chosen 
block that is applied to a set of blocks for classification, 
as shown in Table I, where three different block sizes are 
studied, i.e., 50 × 50, 100 × 100, and 150 × 150. Table I shows 
that the SVM only performs well in the block where it is 
trained. In addition, the block size does affect the classification 
performance. More training samples are available when the 
block size is large, but the cluster overlapping problem in the 
feature space is more significant because more cover types 
may exist in the same block. When the block size is small, 
the overlapping problem is mitigated due to the less cover 
types, while the number of training samples is reduced and 
the generalization performance suffers. There is a trade-off 
between the number of training samples (i.e., the block size) 
and the complexity of feature space. Practically, 100 × 100 
was found to be an appropriate block size in the study area. 
Therefore, Table I validates the effectiveness of the localized 
operation, which assumes independence across blocks.

V. CRP MAPPING IMPLEMENTATION

A. Sample Selection for Training and Evaluation

Considering the error in the existing CRP reference data 
provided by NRCS, we develop a specific way to select 
reliable samples for training and evaluation. Usually, major 
errors in the present CRP reference data are the mis-location 
and/or misalignment of CRP tracts. If this mis-location is 
not significant, we may still get reliable training and testing 
samples by sampling away from CRP boundaries. In other 
words, all data samples are selected from the center areas of 
CRP tracts. A more reliable way to get data sample is the 
field trip, when accurate training samples can be collected 
in CRP tracts. Based on reliable training samples, the CRP 
mapping results can even rectify some locality errors and 
spatial misalignment of CRP tracts in the reference data.

B. CRP Mapping using DTC

In C4.5 DTC, the classification hyperplane consists of a 
set of local splitting operations without guaranteeing global 
optimality. Moreover, the DTC training process often faces 
the overfitting problem, i.e., the learned concept is too specified 
for the training data, which leads to poor generalization 
performance. Some pruning methods have been developed 
to mitigate the overfitting problem, such as an error-based 
pruning (EBP) method suggested in C4.5 [38], [16]. Assume 
there are N training samples covered by a node, E samples 
are misclassified. If this node is pruned, the error rate is 
\( R = E/N \). For given confidence level \( \alpha \), the upper bound 
of the estimated error for the future test can be computed as 
\( R' = R + U_{\alpha}(E, N) \) with the assumption that errors in 
the training set are binomially distributed, where \( U_{\alpha}(E, N) \) is the 
confidence limits for the binomial distribution. This method 
conservatively estimates the misclassification rate when pruned 
trees are applied to the test data.
The EBP method favors higher $P_a$ and $P_b$, while decrease $P_c$, especially when $\alpha$ is small. In this work, based on the same assumption of the EBP method, we develop a recall-based pruning (RBP) approach in favor of higher recall. When splitting a node, the data samples in this node is divided into two parts: $\{a^+, a^-\}$ and $\{b^+, b^-\}$, where $a^+$ is called true positive, $a^-$ is false positive, $b^+$ is false negative, and $b^-$ is true negative. Then the recall of this splitting is defined as:

$$
P_c = \frac{a^+}{a^+ + b^+}. \tag{11}
$$

Different to EBP, RBP begins from the next to the leave layer of DTC because $P_c$ is only associated with those nodes that are not leaves, and the error that needs to be reduced is $b^+$. Therefore, each pruning removes a subtree from the constructed tree. If a subtree is pruned, given confidence level $\alpha$, the upper bound of $b^+$ is estimated as $B^+ = b^+ + U\alpha(b^+, b^+ + b^-)$. Since RBP cannot guarantee small $a^-$, it should be used in conjunction with EBP, i.e., when decide whether to prune a node or not, we compare both $R$ and $B^+$ calculated at this node with those of a subtree.

C. CRP Mapping using SVM

Although the overfitting problem of DTC could be mitigated by pruning, the generalization performance still cannot achieve the optimal solution. Moreover, the curse of dimension could arise if training samples do not significantly outnumber the feature dimension. SVM methods avoid these limitations by optimizing a margin-based criterion, resulting in a better generalization than DTC. However, when deal with the uneven classification problem, SVM usually leads to good $P_a$ but poor $P_c$. This usually happens in text classification [10], [33], [40], as well as CRP mapping [43]. Various relaxation approaches have been developed to address this problem [40]. The principle of these methods is to adjust either or both of the position and orientation of the classification hyperplane to achieve a better performance. We study two relaxation methods in this work. One is the SVM based embedded relaxation (SVM-ER) method that assigns uneven cost to the misclassification of positive and negative samples during the SVM training [44], [36], leading to the change of both position and orientation of the hyperplane. The other is an efficient SVM based post-learning relaxation (SVM-PLR) approach suggested in [40], where an adaptive beta-gamma filtering method [45] is used to adjust the position of the hyperplane.

The SVMlight outputs indicate both the distance of each sample to the decision hyperplane and the class type with the appropriate sign. After ranking these distances from the positive to the negative, we can build a utility model, i.e., $U$, by assigning equal or various weights $w_1$ and $w_2$ to the true positive and false positive according to the class label of training data:

$$
U = w_1 a^+ - w_2 a^- \tag{12}
$$

where $a^+$ and $a^-$ are defined in Section V-B. In this work, we set $w_1 = w_2 = 1$. Based on the utility model, we search for the distance threshold that has the maximum $U$, denoted by $\theta_{opt}$, as well as the threshold of the first zero $U$, called $\theta_{zero}$. Then the final decision threshold is calculated as:

$$
\hat{\theta} = \pi \theta_{zero} + (1 - \pi) \theta_{opt}, \tag{13}
$$

$$
\pi = \beta + (1 - \beta) e^{-N\gamma}, \tag{14}
$$

where $N$ denotes the number of positive class training samples, and $\beta$ and $\gamma$ determine the extent of threshold relaxation from its optimal value. They can be determined by cross validation or experience [24]. Furthermore, given the training data, we want to study how SVM and RBF kernel parameters affect $P_a$, $P_b$ and $P_c$ via the $\xi\alpha$ – estimator suggested in [28]. The $\xi\alpha$ – estimator is a highly efficient approximation to the time-consuming Leave-one-out (LOO) estimator proposed in [34]. Given training data, the LOO estimator can provide an nearly unbiased estimation of the true generalization performance, and the $\xi\alpha$ – estimator provides lower bounds of $P_a$, $P_b$ and $P_c$, which is more conservative than the LOO estimator.
VI. SIMULATION RESULTS

In this section, we investigate the CRP mapping performance in the study area as shown in Fig. 1. After removing CRP boundary areas, remained 60% CRP area is considered as reliable CRP sites, where training and testing samples for data classifier will be selected and used, respectively. Given sampling rate $x$, the equivalent sampling rate (ESR) for CRP is computed as $0.6 \times x$. For example, if $x = 1/3$, the ESR is about 20%, and if $x = 1/6$, the ESR for CRP is about 10%. The selection of non-CRP training samples is done in the same way. CRP mapping is studied based on the localized operation, where the study area is split into blocks of size $100 \times 100$.

A. Simulation of DTC

When C4.5 DTC is used for CRP mapping, the confidence level $\alpha$ is set as 0.01. We first compare the DTC that is not pruned, the one pruned using EBP, and the one pruned using RBP. Given a certain sampling rate, we first select CRP and non-CRP samples for DTC training, and remained samples for testing. The numerical results via cross validation are listed in Table II with two different sampling rates.

<table>
<thead>
<tr>
<th>Sampling Rate</th>
<th>0.2</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>$P_a$</td>
<td>97.07</td>
<td>97.43</td>
</tr>
<tr>
<td></td>
<td>96.14</td>
<td>96.81</td>
</tr>
<tr>
<td>$P_b$</td>
<td>75.22</td>
<td>79.25</td>
</tr>
<tr>
<td></td>
<td>71.19</td>
<td>78.09</td>
</tr>
<tr>
<td>$P_c$</td>
<td>87.27</td>
<td>86.21</td>
</tr>
<tr>
<td></td>
<td>84.88</td>
<td>82.05</td>
</tr>
</tbody>
</table>

From Table II, we can see that when $\alpha = 0.01$, EBP results in higher $P_b$, while $P_c$ is decreased. After using RBP in conjunction with EBP, $P_c$ can be increased with the sacrifice of $P_b$. This is not unexpected because when we try to increase the recall rate ($P_a$), it is unavoidable to mis-classify some non-CRP samples as CRP ones. The tradeoff between $P_b$ and $P_c$ can be predicted via equation (7). For example, in the training and testing data, CRP samples count about 8% over all samples. Therefore, at sampling rate 20%, given $\lambda = 0.08$, $P_a = 97.23\%$, and $P_c = 87.78\%$, we predict that $P_b = 79.41\%$ according to (7), which is close to the true value, i.e., 76.47%. In addition, in order to remove isolated misclassified pixels, we can use the Bayesian context fusion or morphological operation to improve the mapping performance by cleaning up the noisy segmentation map.

We also study the individual contribution of multisource geospatial data to the mapping performance. Given sampling rate 20%, simulation results are shown in Table III, where the numbers in parentheses are the increases compared with the mapping result using the satellite imagery (layer A in Fig. 2) only. It is shown that all multisource data can improve classification performance in terms of $P_a$, $P_b$, and/or $P_c$. (1) Derived features (Layer B) are helpful to discriminate healthy green vegetation from dead vegetation, bare soil and urban areas, but provide limited disparity information among different green vegetation. Both CRP and non-CRP areas consist of multiple vegetation types, therefore Layer B provides slight improvement. (2) From LULC GAP data we know that more than half of this region are covered by crops, which usually show relatively smooth texture behavior, while CRP areas are unmanaged areas covered by different grass species that tends to show less smooth texture behavior. The texture smoothness/roughness can be efficiently captured by a window-based local mean and variance (Layer C), which contribute more to classification accuracy than layer B. (3) The improvement from GIS data (Layer D) is most significant even there are only three GIS layers. This indicates that GIS data has certain correlations with CRP tracts with respect to elevation, distance-to-waterbody and slope. This observation is consistent with the CRP enrollment policy of FSA, justifying the usefulness of multisource GIS data for CRP mapping.

B. Simulation of SVM

There are two free parameters in SVM that need to be determined, i.e., $\sigma$ in RBF kernel and regularization factor $C$, which are usually determined by cross validation or experience. It was found that SVM performs well when $C$ is between 10 to 1000, and $\sigma$ significantly affects precision ($P_b$) and recall ($P_c$). We need to estimate an appropriate $\sigma$ value that leads to high $P_b$ with acceptable $P_c$. Therefore, given the training data, $\xi\alpha - estimator can be used to select proper $\sigma$ by plotting $P_a$, $P_b$, and $P_c$ against $\sigma$ in a certain range, as shown in Fig 5. As we can see, when $\sigma$ is increased from a small value (e.g., 0.1), $P_a$ can achieve its maximum, then gradually decreases. $P_b$ and $P_c$ vary in opposite directions when $\sigma$ is small, which verifies the existence of a tradeoff between them if $P_a$ keeps constant. Then both $P_b$ and $P_c$ decrease after $\sigma = 1.71$ where

\[\xi\alpha = \frac{1}{2\sigma^2} \text{atan}(\frac{-1}{\sigma^2}) - \frac{1}{2\sigma^2}\]

\[\xi\alpha = \frac{1}{2\sigma^2} \text{atan}(\frac{-1}{\sigma^2}) - \frac{1}{2\sigma^2}\]
PER can slightly improve its a preferred value. However, as shown in Table IV, SVM-ER samples in the cost function based on the training data, so that determine a proper relative weight (RW) of CRP and non-CRP. When implementing SVM-ER, we first use \( \xi^\alpha \) high, significant improvement of them could be very difficult. \( P_b \) in a block are classified. Since the original (b) is the mapping results using SVM where all data samples shows the original CRP tracts in the reference data that are two clips in the study area are illustrated in Fig. 6. Fig. 6 (a) superimposes the original CRP, \( \sigma \).

It is shown that both \( P_b \) and \( P_c \) are above 90\% without the relaxation. At sampling rate 10\%, the mapping results of two clips in the study area are illustrated in Fig. 6. Fig. 6 (a) shows the original CRP tracts in the reference data that are considered well aligned with the satellite image, and Fig. 6 (b) is the mapping results using SVM where all data samples in a block are classified. Since the original \( P_b \) and \( P_c \) are very high, significant improvement of them could be very difficult. When implementing SVM-ER, we first use \( \xi^\alpha \) – estimator to determine a proper relative weight (RW) of CRP and non-CRP samples in the cost function based on the training data, so that \( P_c \) could be maximized. In practice, we found that RW=0.5 is a preferred value. However, as shown in Table IV, SVM-ER can slightly improve \( P_b \) and \( P_c \). SVM-PLR can increase \( P_c \) considerably, but \( P_b \) usually suffers. As mentioned before, \( P_b \) can also be estimated by equation (7). For instance, at sampling rate 20\%, when \( P_a = 98.3\% \) and \( P_c = 96.43\% \), we have \( P_b = 84.3\% \) near to the true value, i.e., 82.49\%.

The individual contribution from multisource geospatial data is also studied as shown in Table V at sampling rate 20\%. The simulation results in Tables III and V demonstrate that \( C4.5 \) and SVM are consistent regarding the feature contribution, where texture information and GIS data are the most important features to improve CRP mapping accuracy. It is also shown that SVM works better than DTC under the same sampling rate. This demonstrates that SVM has better generalization performance than DTC.

### Table IV

**Classification Performance of SVM at Different Sampling Rates (I: No Relaxation, II: SVM-ER, III: SVM-PLR)**

<table>
<thead>
<tr>
<th>Sampling Rate</th>
<th>0.2</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>( P_a )</td>
<td>99.10</td>
<td>99.12</td>
</tr>
<tr>
<td>( P_b )</td>
<td>93.01</td>
<td>93.24</td>
</tr>
<tr>
<td>( P_c )</td>
<td>94.41</td>
<td>94.42</td>
</tr>
</tbody>
</table>

We also study the prediction error of SVM via LOO and \( \xi^\alpha \) – estimators. At sampling rate 10\%, different training and testing samples are selected to estimate the mean and standard deviation of the prediction error. Simulation results are shown as error bars in Fig. 7. In the study area, there are 19 out of 25 blocks which have significant CRP tracts. The dash and dot lines indicate LOO and \( \xi^\alpha \) estimations, respectively. Since \( \xi^\alpha \) – estimator provides lower bounds of the estimation, the prediction is more conservative but more efficient than LOO estimation. Both of them can be used to predict the performance of CRP mapping. Furthermore, they could also be used to measure the effectiveness of given training samples. If predicted errors are significant, we might want to select more representative training samples or add more training samples.

### C. Rectification of CRP Reference Data

In the case of CRP mapping, we want to use as few as possible training samples, since the CRP reference data might not be available and training samples have to be acquired from field trips. On the other hand, when we have inaccurate CRP reference data, we may want to rectify it using the proposed SVM-based approach. In this case, more reliable training samples are preferred. Therefore, in the study area, all remained 60\% CRP areas (by removing boundaries) are used for SVM training, and we can regroup classification results from all blocks to produce a rectified CRP map. Two examples are shown in Fig. 8. Fig. 8 (a) superimposes the original CRP reference data on the Landsat image in band 4, and Fig. 8 (b) shows the rectification results. It is shown the mis-locality and misalignment errors can be significantly reduced around CRP boundaries. Since about 40\% CRP boundary areas are not used for SVM training, better rectification results could be obtained by selecting more training samples near boundaries or using bootstrap approaches where the classification result is used to train a new SVM for data classification.
generalization performance than DTC in this work. Our future source geospatial data is manifested by their improvements in precision, while SVM-PLR can enhance recall significantly, while SVM-ER cannot improve recall rate. Two relaxation methods are also studied for SVM to improve recall, while SVM-ER cannot improve recall significantly, while SVM-PLR can enhance recall with acceptable precision if we properly choose the relaxation parameters. In addition, the individual contribution of multisource geospatial data is manifested by their improvements on CRP mapping accuracy. Overall, SVM shows a better generalization performance than DTC in this work. Our future research will focus on CRP compliance monitoring based on the proposed CRP mapping approaches.

![Fig. 7. Predication errors of LOO and ξα – estimators. (a) Classification accuracy (Pa). (b) Precision (Pp) (c) Recall (Pc).](image)

![Fig. 8. CRP rectification results (145 × 145 pixels): (a) Original CRP reference data. (b) Rectified CRP map](image)

**VII. Conclusions**

We have studied the implementation of DTC and SVM for automatic CRP mapping and rectification, which is complex classification problem due to the complicity of feature space. Particularly, a parallel localized classification framework is suggested and validated based on a study area. Considering the importance of classification sensitivity, a new DTC pruning method is proposed to enhance the recall rate. Two relaxation methods are also studied for SVM to improve recall. Simulation results indicates that SVM-ER cannot improve recall significantly, while SVM-PLR can enhance recall with acceptable precision if we properly choose the relaxation parameters. In addition, the individual contribution of multisource geospatial data is manifested by their improvements on CRP mapping accuracy. Overall, SVM shows a better generalization performance than DTC in this work. Our future research will focus on CRP compliance monitoring based on the proposed CRP mapping approaches.

**ACKNOWLEDGEMENT**

The authors acknowledge the anonymous reviewers for their suggestions and comments which greatly improve this paper.

**REFERENCES**


