Automatic CRP Mapping using Nonparametric Machine Learning Approaches

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

Abstract—This paper studies an uneven two-class unsupervised 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 recent payments of nearly $1.6 billion for new enrollments (2003 signup), it is imperative to obtain accurate digital CRP maps for management and evaluation purposes. CRP mapping is a complex classification problem where both CRP and non-CRP areas are composed of various cover types. Two nonparametric machine learning approaches, i.e., decision tree classifier (DTC) and support vector machine (SVM) are implemented in this work. Specifically, considering the importance of CRP classification sensitivity, a new DTC pruning method is proposed to increase recall. We also study two SVM relaxation approaches to increase recall. Moreover, a localized and parallel framework is suggested in order to efficiently deal with the large scale CRP mapping need. Simulation results validate the applicability of the 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 in past decades. Conventional parametric statistical model-based methods show their efficiency in such problems [1], [2]. In recent years, various multisource geospatial data is used to facilitate the classification of multispectral imagery [3], [4], [5], [6]. Correspondingly, people found that it may not be appropriate to model multisource data by traditional multivariate statistical models [3], [4], [7], [8], [9]. Therefore, nonparametric methods should be considered. In this work, we study nonparametric machine learning approaches for mapping United States Department of Agriculture (USDA)’s Conservation 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 [10]. Very little work has been done for CRP mapping so far, and recent work in [11] requires considerable human interpretation and intervention.

This work is supported by Oklahoma NASA EPSCoR Research Initiation Grant (2003) and Water Research Center Grant (2003, 2004), Oklahoma State University.

X. Song and G. Fan are with the School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA (e-mail: {xiaom.song, gfan}@okstate.edu).

M. Rao is with the Department of Geography, Oklahoma State University, Stillwater, OK 74078, USA (e-mail: mahesh@okstate.edu).

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 provided by Natural Resources Conservation Service (NRCS) is not very accurate or up-to-date, and we need a specific way to select reliable training samples and to evaluate the mapping performance from the present reference data. Moreover, based on the mapping results, we can also correct some errors in the present reference data. (3) CRP mapping is an uneven classification task where CRP tracts usually consist of less than 10% over all study areas. Accordingly, methods that favor high recall rates should be considered. (4) Since CRP mapping 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 multisource GIS data is prominent in the CRP mapping application. For example, if a significant 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 [6] shows the advantages of using neural networks for the classification of complex rural areas. In this work, we study the decision tree classifier (DTC) and support vector machine (SVM) [12] for CRP mapping. The principle of the DTC is to break up a complex classification problem into a union of several simpler classification issues. The SVM constructs a linear classification hyperplane that maximizes the margin between two different training patterns in the original feature space or a high dimensional feature space generated by kernel methods [12], [13], [14], [15]. In our recent work [16], both the 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, there is usually no significant mis-location of CRP tracts in the reference data; therefore, 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 classifier training and to evaluate the performance of CRP mapping.

Since the CRP mapping is an uneven 2-class unsuper-
vized 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. In our previous work [16], we found that pruned DTCs and SVMs 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 problems related to CRP mapping, e.g., 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 the SVM to improve recall specifically. Moreover, we propose a localized and parallel classification framework to implement CRP mapping for large areas efficiently and effectively.

This paper is organized as follows. In Section II, USDA’s CRP program and the study area will be introduced. The DTC and the 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 the DTC and the 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. The CRP is a provision of the 1985 Farm Bill that seeks to convert highly erodible lands with active crop production to permanent vegetative cover [17]. 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 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 farms get selected for 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.

Currently, there is no standardized approach to keeping track of existing CRP tracts. FSA relies on aerial photography to manually delineate CRP tracts on a county level basis. These aerial photographs are at the section level, and provide little information about the CRP from a landscape perspective. Furthermore, when NRCS drafts the CRP reference data, the possible mis-locality and spatial misalignment of CRP tracts deteriorate the reliability and usability of these reference data [11]. Therefore, the goal of this work is to develop an automatic tool for accurate CRP mapping 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 260 km², where the accurate CRP mapping framework is being 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 salient region for the study of CRP mapping.

III MACHINE LEARNING APPROACHES

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 to acquire 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 their learning goal is to determine a hypothesis of the target concept that best fits the training data. The DTC has shown advantages in real remote sensing (RS) applications for more than ten years [4], [18], [19], [20], [21], [22]; however, considering that the overfitting problem is met by the DTC with poor generalization performance, the SVM is suggested as an alternative to the DTC. Recent research on the SVM in RS applications have shown impressive classification results [23], [24], [25], [26], [27], [28], [29]. In this work, both the DTC and SVM are used for CRP mapping as a semi-supervised classification issue involving multisource geospatial data. Particularly, the generalization performance of the two machine learning approaches is carefully studied to produce accurate CPR maps with high recall rates.
III-A Decision Tree Classifier (DTC)

The 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 CRP mapping [30]. 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 [31]. 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 this work, the tree stops growing if there are less than five samples in a node.

III-B Support Vector Machine (SVM)

SVMs are newly developed learning methods [12]. 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 a classification hyperplane induced from the training samples that maximally separates classes, or equivalently, to minimize $\frac{1}{2}\|w\|^2$, 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 $\xi_i \geq 0$, $i = 1, \ldots, l$ are 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 [13], [15]. There are several often used kernel functions, such as the radial basis function (RBF):

$$K(x, x_i) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right),$$

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 [32] is used to perform training and classification of geospatial database. SVM parameters, i.e., $\sigma$ in RBF kernel and regularization factor $C$, are usually determined by cross validation or experience [33], [34].

III-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 ($P_a$) is defined as the percentage of pixels that are correctly classified in terms of CRP and non-CRP. Precision ($P_c$) indicates the percentage of detected CRP pixels that are true ones. Recall ($P_r$) is the percentage of true CRP pixels that can be detected. 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 [35]. 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 [35], a tradeoff between precision and recall is formulated as:

$$\lambda P_c + (\lambda + P_a - 1)P_r = 2\lambda P_k P_c,$$

where $\lambda$ 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, it is necessary 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

IVA Geospatial Database

The geospatial database is composed of the Landsat TM satellite imagery, vegetation indices, texture information, and GIS data, which are all in raster format. Combined with LULC GAP data and CRP reference data, there are a total of 40 layers as shown in Fig. 2. During CRP mapping, layer sets A and D are original inputs, and layer sets B and C are automatically generated by the system during run time.

Layer set A consists of the Landsat TM multispectral images obtained in February and June of year 2000 with a resolution of $30m \times 30m$. Since Band 1 is prone to scattering, we do not use it in the database. 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. All layers were geometrically and radiometrically corrected before being applied to the classification. The accuracy of the geometric correction is 0.5 pixels, and the technique outlined in [36] is used to perform radiometric correction.

Layer set B contains vegetation indices that include the Normalized Vegetation Difference Index (NDVI), and band ratios TM4/TM3, TM5/TM2, TM5/TM4. The NDVI for Landsat TM is computed as:

\[
NDVI = \frac{TM4 - TM3}{TM4 + TM3},
\]

where TM3 and TM4 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. Band ratio TM4/TM3 (Ratio vegetation index) is widely used for vegetation discrimination. Ratio TM5/TM2 is helpful to discriminate different vegetation types [37]. TM5/TM4 (Ratio drought index) can provide more information of plant water content [38], which is useful to discriminate irrigated crops from relatively dry CRP grasses. Totally, there are 7 layers in Layer set B that are composed of NDVI (1 layer), and three band ratios of two seasons (6 layers).

Layer set C consists of 20 layers of texture information, including local mean and local variance of each band in each season. The local mean and local variance are computed on the spectral value within a 3 × 3 window. The texture layers are followed by GIS data of Layer set 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.

**IV-B Feature Extraction**

The geospatial database is directly applied to the DTC that generates a set of rules that are easy to interpret and understand. On the other hand, the database needs to be preprocessed before implementing SVM. First, it is necessary to normalize each data layer to be zero-mean and unit variance. This normalization can balance the relative importance between different layers. Second, since the Landsat TM spectral channels and the derived features contain highly redundant information, it is necessary to reduce the feature redundancy via feature selection or extraction. In this work, we use discriminant analysis feature extraction (DAFE) [39] to extract feature subsets from 5 multispectral image bands, 5 layers of local mean, as well as 5 layers of local variance for each season separately. In each set of 5-layer data, the three layers of extracted features with the largest eigenvalues are preserved. As the result, there are 9 layers for each season and totally 18 layers for two seasons. Including 7 layers of vegetation indices and 3 GIS layers, there are totally 28 layers for SVM-based CRP mapping. Since DAFE only works well when CRP and non-CRP are normally distributed, it might not be able to produce the most discriminative features. It is expected that more effective features could be obtained by using advanced feature extraction methods [39], [40].

**IV-C Localized Data Classification**

When selecting training samples, a straightforward way is to select samples from the whole study area. Nevertheless, it is worth pointing out that there are more than 30 different CRP species in Texas County, Oklahoma, and there are 14 CRP species in the study area shown in Fig. 1. On the other hand, there exist many other cover types in non-CRP regions as well, such as crop, urban, and pasture, etc. For example, the 3-D spectral feature distributions of both CRP and non-CRP in the study area are illustrated in Fig. 3 (a). It can be easily observed that the spectral features of CRP and non-CRP highly overlap in bands 3, 4, and 7.

In this work, we suggest a localized block-based technique to achieve automatic CRP mapping efficiently. The proposed technique splits the study area into small blocks, and the DTC or SVM training and classification are performed within each block independently. Then the outputs of all blocks are combined to rebuild the whole CRP map. The classification framework is shown in Fig. 4.
to a parallel classification structure. For example, when 20% of CRP and non-CRP areas are used for training and the remaining areas are used for testing, it takes more than half an hour (Pentium IV 2.2GHz CPU, 1GB memory) on the whole study area. If the study area is split into 25 blocks of size around $100 \times 100$ pixels, the time for SVM training and classification in each block is about 10 seconds, and 30 seconds for the DTC. If we have a parallel computing architecture, the whole area can be processed efficiently.

Figure 3. (a) 3-D feature spaces of CRP and non-CRP regions. (b) Overlap of CRP species in 3-D feature spaces, where species type 1 is Old World Bluestem, type 2 is Plains Bluestem, type 3 is WW Spar, type 4 is Ganada, type 5 is Plains Bluestem (1986), type 6 is Ganada (1986), type 7 is Old World Bluestem (1987), type 8 is Caucasian (1987), type 9 is Plains Bluestem (1987), type 10 is Plains (1988), type 11 is Plains (1989), type 12 is WW Spar (1989), type 13 is Old World Bluestem (1990), and type 14 is Native Mixture (1990).

The proposed block-based technique assumes independence across all local blocks. In reality, each block is not completely independent to other blocks. We manifest this fact by randomly selecting five blocks. Three different block sizes are studied: $50 \times 50$, $100 \times 100$, and $150 \times 150$ pixels. The SVM is trained from each of five blocks, and it is used to classify all five blocks. Simulation results are listed in Table I, where NA denotes no sample is detected as CRP. It is shown that a trained SVM only performs well in the block where it is trained. Since CRP mapping is an uneven classification problem, $P_a = 90\%$ does not mean a good performance without high $P_b$ and $P_c$. As we can see, when the block size is smaller, the CRP and non-CRP cover types in one block tend to be purer and more distinct compared with other blocks. Then the locally trained SVM may not be applicable to other areas. Moreover, if block size is larger, due to more complex cover types in each block, the applicability of a locally trained SVM to other areas is not good yet. Therefore, this result validates the assumption of block independence.

CRP training samples indicated by the reference data are usually available for each block if the block size is large enough. However, the block-based classification might be limited in the case that there is an unknown block without training samples. This is due to the fact that a classifier trained in one block is usually not applicable to others. In order to deal with this problem, we suggest two SVM-based approaches. One is to use multiple one-class SVMs (OCSVM) [41], [42]. OCSVM is an extension of SVM that constructs a decision hyperplane or hypersphere to separate majority data samples from outliers in the feature space. Particularly, each OCSVM is trained based on a single CRP species only, and applied to detect if there exists the same species in the unknown block. The final CRP map of the unknown block is obtained by merging all OCVSM outcomes, each of which is corresponding to one CRP species. This method is efficient due to the fact that we only consider the CRP data. The other approach to classify the unknown block is the “one-against-all” multi-class SVM method [43], [44] with multiple SVMs, each of which is trained with samples from a single CRP species as the positive class, and with samples of all other CRP and non-CRP cover types as the negative class. The effectiveness of this method has already been shown in land cover classification problems [45], [46].

V CRP MAPPING IMPLEMENTATION

V-A Sample Selection for Training and Evaluation

Considering the error in the existing CRP reference data provided by NRCS, we develop a specific method to select reliable samples for classifier training and evaluation. The majority of errors in the present CRP reference data are the mis-location of CRP tracts and/or CRP boundaries. If this mis-location is not significant (usually true in most cases), 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 samples is to perform field study of the CRP tracts in question. Based on reliable training samples, the CRP mapping results can even correct some locality errors and spatial misalignment of CRP tracts in the reference data.

V-B CRP Mapping using DTC

In the 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. Therefore, some pruning methods have been developed to mitigate the overfitting problem. We use a post-pruning approach suggested in C4.5 [30], which is also called error-based pruning (EBP) [47]. For this approach, we assume...
<table>
<thead>
<tr>
<th>TABLE I</th>
<th>THE STUDY OF INTER-BLOCK DEPENDENCY VIA TRAINING-CLASSIFICATION PROCESS AT 20% SAMPLING RATE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 x 50</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>1</td>
<td>96.92</td>
</tr>
<tr>
<td>2</td>
<td>95.42</td>
</tr>
<tr>
<td>3</td>
<td>96.88</td>
</tr>
<tr>
<td>4</td>
<td>97.22</td>
</tr>
<tr>
<td>5</td>
<td>95.29</td>
</tr>
<tr>
<td>100 x 100</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>96.96</td>
</tr>
<tr>
<td>2</td>
<td>96.24</td>
</tr>
<tr>
<td>3</td>
<td>96.10</td>
</tr>
<tr>
<td>4</td>
<td>96.57</td>
</tr>
<tr>
<td>5</td>
<td>96.57</td>
</tr>
<tr>
<td>150 x 150</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>96.56</td>
</tr>
<tr>
<td>2</td>
<td>95.18</td>
</tr>
<tr>
<td>3</td>
<td>97.16</td>
</tr>
<tr>
<td>4</td>
<td>96.11</td>
</tr>
<tr>
<td>5</td>
<td>95.18</td>
</tr>
</tbody>
</table>

There are $N$ training samples covered by a node and $E$ samples are misclassified. If this node is pruned, the error rate is $R = E/N$. For a 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 limit for the binomial distribution. This method conservatively estimates the misclassification rate when pruned trees are applied to the test data.

The EBP method favors $P_a$ and $P_b$, while decreasing $P_c$, 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 are 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:

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

(9)

Contrast to EBP, RBP begins from the parent node of each leaf node in the 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 deciding whether to prune a node or not, we compare both $R'$ and $B^+$ calculated at the current node with those nodes of a subtree.

V-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 dealing with the uneven classification problem, SVM usually leads to good $P_b$ but poor $P_c$. This usually happens in text classification [48], [49], [50], as well as in CRP mapping [16]. Various relaxation approaches have been developed to address this problem [50]. 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 costs to the misclassification of positive and negative samples during the SVM training [12], [51], 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 [50], where an adaptive beta-gamma filtering method [52] 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 $U$ 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:

$$U = w_1a^+ - w_2a^-,$$  

(10)

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:

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

(11)

$$\pi = b + (1 - b)e^{-NT},$$  

(12)

where $N$ denotes the number of positive class training sam-
samples, and \( \beta \) and \( \gamma \) determine the extent of threshold relaxation from the threshold’s optimal value. \( \beta \) and \( \gamma \) can be determined by cross validation or experience [53]. 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 [54]. The \( \xi \alpha - \) estimator is a highly efficient approximation to the time-consuming Leave-one-out (LOO) estimator proposed in [55]. 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, the remaining 60% of the CRP area is considered as reliable CRP sites, where training and testing samples for the data classifier will be selected and used, respectively. Given a sampling rate \( x \), the equivalent sampling rate (ESR) for CRP areas is computed as \( 0.6 : x \). 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 block-based operation, where the block size is around 100 \( \times \) 100 pixels.

VI-A Simulation of DTC

When the C4.5 DTC is used for CRP mapping, the confidence level \( \alpha \) is set as 0.05. We first compare the DTC that is not pruned, the one pruned using EBP, and the one pruned using RBP. Given a sampling rate, we first select CRP and non-CRP samples for DTC training, and we use the remaining samples for testing \(^1\). The numerical results via cross validation are listed in Table II with two different ESRs.

<table>
<thead>
<tr>
<th>ESR</th>
<th>20%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_a )</td>
<td>97.12</td>
<td>97.56</td>
</tr>
<tr>
<td>( P_b )</td>
<td>75.72</td>
<td>80.29</td>
</tr>
<tr>
<td>( P_c )</td>
<td>87.17</td>
<td>86.92</td>
</tr>
</tbody>
</table>

\(^1\) Removed CRP regions are excluded in both training and numerical testing.

From Table II, we can see that when \( \alpha = 0.05 \), 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 expected because when we try to increase the recall rate \( P_c \), misclassification of non-CRP samples as CRP samples will occur. The tradeoff between \( P_b \) and \( P_c \) can be predicted via equation (7). For example, in the training and testing data, CRP samples consist of about 8% of all samples. Therefore, at ESR 20%, given \( \lambda = 0.08 \), \( P_a = 97.37% \), and \( P_c = 87.68\% \), we predict that \( P_b = 80.84\% \) according to (7), which is close to the true value, i.e., 77.93%. Moreover, we can further improve the mapping performance by using Bayesian context fusion or morphological operation to remove the isolated misclassified pixels.

We also study the contributions from different combinations of multisource data to the mapping performance. Given 20% ESR, 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 set A in Fig. 2) only. It is shown that all multisource data can improve classification performance in terms of \( P_a \), \( P_b \), and \( P_c \). (1) Vegetation indices (Layer set B) provide helpful information to discriminate healthy green vegetation from dead vegetation, bare soil, and urban areas, as well as limited disparity information among different green vegetation. The difficult part of CRP mapping is the discrimination of different vegetation types, and layer set B provides only slight improvements. (2) From LULC GAP data we know that more than half of this region is covered by crops, which usually show relatively smooth texture behavior, while CRP areas are unmanaged areas covered by different grass species that tend to show less smooth texture behavior. The texture smoothness/roughness can be efficiently captured by a window-based local mean and variance (Layer set C), which contribute more to classification accuracy than layer set B. (3) The improvement from GIS data (Layer set D) is most significant when 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.

VI-B Simulation of SVM

From the cross validation, it was found that SVM performs well when \( C \) is between 10 to 1000, while \( \sigma \) significantly affects precision \( P_b \) and recall \( P_c \). We need to estimate an appropriate \( \sigma \) value that leads to high \( P_c \) with acceptable \( P_b \). Therefore, given the training data, the \( \xi \alpha - \) estimator can be used to select a 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, \( P_a \) varies slightly. \( P_b \) and \( P_c \) vary in opposite directions when \( \sigma \) is small, which verifies the existence of a tradeoff between them, if \( P_a \) remains approximately constant. Then both \( P_b \) and \( P_c \) decrease after \( \sigma = 2.13 \) where \( P_b = 92\% \) and \( P_c = 88.05\% \). At this point, \( P_c \) achieves its highest lower bound. Considering
the importance of high $P_c$, we set $\sigma = 2.13$ in this work. Specifically, we use two relaxation methods introduced in Section V-C to increase $P_c$. Simulation results using cross validation are listed in Table IV at two different ESRs.

**TABLE IV**

<table>
<thead>
<tr>
<th>Classification performance of SVM at different ESRs (I: NO RELAXATION, II: SVM-ER, III: SVM-PLR )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESR</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$P_a$</td>
</tr>
<tr>
<td>$P_b$</td>
</tr>
<tr>
<td>$P_c$</td>
</tr>
</tbody>
</table>

It is shown above that both $P_b$ and $P_c$ are more than 90% without the relaxation. At 10% ESR, the mapping results of four clips in the study area are illustrated in Fig. 6. Fig. 6 (a) shows the original CRP tracts in the reference data, and Fig. 6 (b) depicts the mapping results using SVM where all data samples in a block are classified. Since the original $P_b$ and $P_c$ are quite high, significant improvement of them could be very difficult. When implementing SVM-ER, we first use the $\xi\alpha$ 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. We found that RW=0.5 is a preferred value. However, as shown in Table IV, SVM-ER can slightly improve $P_b$ and/or $P_c$. 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 20% ESR, when $P_a = 98.47\%$ and $P_c = 96.89\%$, we have $P_b = 85.8\%$ near to the true value, i.e., 83.97\%.

The contributions from different combinations of multi-source data are also studied and listed in Table V at 20% ESR. The simulation results in Tables III and V demonstrate that the C4.5 and SVM are consistent regarding the feature contribution, where texture information and GIS data are the most important features used 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.

We also study the prediction error of SVM via LOO and $\xi\alpha$ estimations, respectively. In the study area, there are 19 out of 25 blocks having significant CRP tracts. Since $\xi\alpha$ estimator provides lower bounds of the estimation, the prediction is more conservative but more efficient than LOO estimation. Both estimators can be used to predict the CRP mapping performance. Furthermore, the estimators can also measure the effectiveness and representativeness of training samples. If predicted errors are significant, we could add more representative training samples.

**VII CONCLUSIONS**

We have studied the application of DTC and SVM for automatic CRP mapping, which is a classification problem of complex rural areas. 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 LOO and $\xi\alpha$ estimations, respectively.
SVM to improve recall. Simulation results indicate 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 its 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, and Mr. David Monismith for their detailed comments that greatly improve this paper.

REFERENCES


