

Eyes in the Skies: A Data-Driven Fusion Approach to Identifying Drug Crops From Remote Sensing Images

Anselmo Ferreira^{ID}, Siovani C. Felipussi, Ramon Pires^{ID}, Sandra Avila, Geise Santos, Jorge Lambert,
Jiwu Huang^{ID}, and Anderson Rocha^{ID}

Abstract—Automatic classification of sensitive content in remote sensing images, such as drug crop sites, is a promising task, as it can aid law-enforcement institutions in fighting illegal drug dealers worldwide, while, at the same time, it can help monitor legalized crops in countries that regulate them. However, existing art on detecting drug crops from remote sensing images is limited in some key factors, not taking full advantage of the available hyperspectral information for analysis. In this paper, departing from these methods, we propose a data-driven ensemble method to detect drug sites from remote sensing images. Our method comprises different convolutional neural network architectures applied to distinct image representations, which are able to represent complementary characterizations of such crops. To validate the proposed approach, we considered in our experiments a dataset containing *Cannabis Sativa* crops, spotted by police operations in a Brazilian region called the *Marijuana Polygon*. The results in this dataset show that our ensemble approach outperforms other data-driven and feature-engineering methods in a real-world experimental setup, in which unbalanced samples are present and acquisitions from different places in the same region are used for training and testing the methods, highlighting the promising use of this solution to

aid police operations in detecting and collecting evidence of such sensitive content properly.

Index Terms—Convolutional neural networks (CNNs), deep learning, detection of drug crops, sensitive remote sensing analysis.

I. INTRODUCTION

SEVERAL studies demonstrate how the use of illicit drugs, such as *Diamorphine* (a.k.a. Heroin) and *Cannabis Sativa* (a.k.a. Marijuana), severely affects the human body [1]–[3]. In the specific case of *Cannabis Sativa*, the ease of manufacturing of such drug makes it the most consumed drug worldwide, being cultivated in more than 130 countries [4]. Unlike other drugs, it does not require specific weather conditions for cultivation. This makes the inspection by security forces of countries that forbid this drug's use a difficult task, as any place can be a potential producer. According to the 2009 United Nations Office of Drugs and Crime World Drug Report [5], the total area of such crops was estimated between 200 000 and 406 000 hectares, which highlights the magnitude of the problem of monitoring such amount of land cover. To minimize such problem, some governments are controlling the commerce of such drug, regulating places where it can be consumed and even the land area allowed for its cultivation [6]. Hence, a solution that aids the remote monitoring of such crops is paramount, as it could be used to automatically acquire evidence in countries that forbid these crops and also to control plantation areas in countries that regulate them.

The use of image processing techniques to help identify sensitive content in images has been touched by the scientific literature in recent years. Examples of such applications include detecting nudity [7], [8], child pornography [9], [10], covert photos [11], [12], among others. Leveraging these approaches in a real-world scenario could aid law-enforcement agencies in collecting crime evidence more effectively and more accurately, as thousands of images could be investigated in a short period with reduced human intervention. However, even with a wide range of tools that make available remote sensing images with low associated costs or even for free (such as Google Maps and Baidu Maps), limited effort has been spent by the scientific literature on detecting sensitive remote sensing data such as drug sites. Although there are some applications proposed to detect digital images containing examples of *Cannabis Sativa* in websites

Manuscript received April 21, 2019; accepted May 5, 2019. This work was supported in part by the National Science Foundation of China under Grants 61332012 and U1636202, in part by the Shenzhen Research and Development Program under Grant JCYJ20160328144421330, in part by the Brazilian Coordination for the Improvement of Higher Level Education Personnel under Grant #99999.002341/2015-08 and the DeepEyes project, in part by the FINEP National Defense and Public Security Grant, in part by the Brazilian National Council for Scientific and Technological Development under Grants #477662/2013-7, #304352/2012-8, #449638/2014-6, and #304472/2015-8, and in part by the São Paulo Research Foundation – FAPESP under Grant #2017/12646-3, DéjàVu project, and in part by Microsoft Research. (*Corresponding author:* Anselmo Ferreira.)

A. Ferreira and J. Huang are with the Guangdong Key Laboratory of Intelligent Information Processing, the Shenzhen Key Laboratory of Media Security, College of Information Engineering, and the National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, China (e-mail: anselmo@szu.edu.cn; jhuang@szu.edu.cn).

S. C. Felipussi is with the Department of Computer Science, Universidade Federal de São Carlos, São Carlos 13565-905, Brazil (e-mail: siovani@hotmail.com).

R. Pires, S. Avila, G. Santos, and A. Rocha are with the Institute of Computing, University of Campinas (Unicamp), São Paulo 13083-852, Brazil (e-mail: pires.ramon@ic.unicamp.br; sandraeliza@gmail.com; geise.kss@gmail.com; anderson.rocha@ic.unicamp.br).

J. Lambert is with the National Institute of Criminology, Brazilian Federal Police, Brasilia 70610-20, Brazil, and also with the Department of Mechatronics Systems, University of Brasilia, Brasilia 70910-900, Brazil (e-mail: lambert.jal@dpf.gov.br).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2019.2917024

[13]–[15], the dataset considered in those applications does not consider remote sensing acquisition, not being useful to perform remote monitoring of such plants covering large areas.

In this paper, we conduct what is, as far as we know, the most complete study on characterizing sensitive remote sensing data containing *Cannabis Sativa* drug crops, by proposing the first data-driven approach focused on identifying such drug crops from remote sensing images. We focus on detecting *Cannabis Sativa* crops as: 1) they are of particular interest in drug policies nowadays—being present in more than 130 countries—covering 92% of the world population [4]; 2) the particular difficulty of classifying such crops, as they can be mixed up with natural vegetation; and 3) the area regulation intention of some countries with respect to this kind of crop, which can create an additional application to new designed methods. The proposed technique comprises an ensemble of convolutional neural networks (CNNs) with different architectures, applied on contrasting input-image representations that are able to highlight specific green areas associated with this crop of interest differently from other Earth materials.

To achieve our goal, we perform an extensive series of experiments in a dataset containing images from a place located in a specific Brazilian region commonly known for containing such crops, called the *Marijuana Polygon*. Experimental results show that the proposed ensemble approach and one of the proposed individual CNNs are the first solutions to cross the 90% classification accuracy mark with low false positive rates, highlighting their promising use in real-world settings in which there are unbalanced samples, photos are taken from different places in the same region and various acquisition conditions can happen.

In short, the main contributions of this paper are as follows.

- 1) We introduce what is, as far as we know, the first public dataset of remote sensing images containing *Cannabis Sativa* crops from real police operations. This dataset was created with the collaboration of the Brazilian Federal Police, which provided information about locations and dates where these crops were found in a Brazilian region referred to as the *Marijuana Polygon*.
- 2) We propose a CNN ensemble tailored to *Cannabis Sativa* drug crops' identification. The proposed approach relies on a CNN that processes the output of different CNN architectures fed with two image representations: the false-color image transform and near-infrared.
- 3) We compare the proposed approach with other 22 methods, which rely on classical feature engineering representations, and five data-driven baselines, posing our work likely as the most complete comparative study on remote sensing drug crop image classification to date.

Finally, for reproducibility purposes, the dataset and the source-code related to this paper are available at IEEE DataPort¹ and Github,² respectively, fostering the development of other approaches focused on drug crop detection using remote sensing imagery.

The remaining of this paper is organized as follows. Section II shows some basic concepts about deep learning through CNNs. Section III shows some related work on remote sensing crops classification, and Section IV presents the dataset we created for this application. Section V presents the proposed approach. Section VI reports the experimental setup used to validate the discussed techniques, while Section VII shows the experimental results. Finally, Section VIII concludes this paper and points out future research directions.

II. BASIC CONCEPTS

Neural networks in which their learnable layers are designed to perform convolution in input signals are called CNNs. The most common operation in CNNs is the convolution, which operates at each (x, y) pixel. Convolutional layers in CNNs are specified by a certain number of filters (or kernels), K , along with additive biases, b , for each filter. These layers operate by computing the convolution of the input-images with each of those filters, adding one bias per output image. Then, activation functions are applied to all the pixels in the output images, also called feature maps. The feature map o , which is the output of a convolution, is calculated as

$$o(x, y) = \sigma((K * I)_{x,y} + b) \quad (1)$$

where σ is the activation function, K the convolution filter, I the input image, and b the bias.

Typical hyperparameters that a CNN architect must think of include: 1) number of filters and bias terms to be used in each convolutional layer (depth); 2) filter size (width and height); 3) filter offset in image pixels to apply the next convolution in the input image (also called stride); and 4) padding factor, to control the spatial sizes of feature maps' volumes.

Loss functions are used in CNNs to measure the errors and to update filter weights in the training step. They work by comparing the ground-truth of the training set and the predicted values of them by the CNN. The most common used loss function, which is also of interest in this paper is the cross-entropy loss, defined as

$$L = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^{|C|} Y_k^{(i)} \log \hat{Y}_k^{(i)} \quad (2)$$

where n is the number of training samples, $|C|$ the number of classes, Y_k the observed label of the training sample, and \hat{Y}_k the predicted label for the training sample.

After defining the loss function, the layer parameters (weights and bias terms) must be found in such a way that they minimize the loss, thereby generating the most precise CNN as possible for classification procedures. Finding these parameters is commonly done using gradient descent methods.

Operations done by convolutional and fully connected layers are not typically meant to be the only operations of a CNN. Thus, several other layers, focusing on diverse auxiliary operations, can also be used and organized differently along with convolutional and fully connected layers, in such a way to generate different CNN architectures suitable for specific applications.

¹<http://dx.doi.org/10.21227/H2WD42>

²<http://www.github.com/anselmoferreira/remote-sensing-sensitive-analysis>

Some of these auxiliary layers are the pooling layers, activation layers, dropout layers, softmax layers, among others.

Different CNN architectures have been proposed in prior art for various problems over the years, using for that several combinations of layers, initialization methodologies, and optimization algorithms, such as the VGG network [16], RESNET [17], INCEPTION (or GoogleLeNet) [18], DENSENET [19], among others. We show later in this paper how such deep networks can be useful for the problem presented herein.

III. RELATED WORK

A wide range of approaches has been proposed in the literature to identify specific crops from remote sensing images. In this section, we subdivide the related work discussion in approaches focused on identifying general crops of interest and approaches focused on detecting drug crops from remote sensing images.

Approaches focused on identifying the crops of interest typically apply the investigation on specific image channels and image transformations. Santoso *et al.* [20] used spatial and morphological filters on false-color images to detect and count oil palm trees. Camps-Valls *et al.* [21] carried out a comparative study of using support vector machines (SVMs) on hyperspectral image pixels to identify different crops, comparing this classifier to neural and fuzzy classification methods. They concluded that SVMs yield better outcomes regarding accuracy, simplicity, and robustness. A similar conclusion was drawn by Khatami *et al.* [22] in a very comprehensive study on the state-of-the-art supervised pixel-based methods for land cover mapping. An interesting tool was proposed by Munoz-Mari *et al.* [23] to validate several algorithms of land cover mapping, by making available 43 remote-sensing image datasets containing several plantation areas. Several other approaches have used transformed pixels as input to classification tools [24], [25], feature selection [26], [27], active learning [28], [29], ensemble of classifiers [30], [31], visual dictionaries [32], [33], and CNNs [34], [35]. More details about common crop-detection approaches can be found in [36].

In the case of detecting drug plantations in remote sensing images, few solutions have been explored to date. For the special case of *Cannabis Sativa* crops, the pioneer work of Daughtry and Walthall [37] analyzed the spectral signatures of such crops. In a later work, Walthall *et al.* [38] proposed a simulation model of this plant to estimate when farmers would be on site, as this crop requires special treatment. In another work, Walthall *et al.* [39] tackled the problem of similar spectral signatures of *Cannabis Sativa* leaves and other plants by proposing spatial signatures, using edge and Fourier analysis of *Cannabis Sativa* and other vegetation leaves. Thiessen [40] showed that, in a specific region in Canada, the *Cannabis Sativa* spectral signature is distinguishable from those of other types of vegetation using IKONOS, Quickbird, and CASI remote sensors. Azaria *et al.* [41] used hyperspectral spectroscopy and feature reduction to find significant spectral signatures of this crop. Finally, the work of Lisita *et al.* [42] relied on region-based semiautomatic classification using multiscale segmentation to identify drug crops.

In closing, we highlight that approaches focused on the identification of drug sites, such as *Cannabis Sativa* plantation areas, are often limited because: 1) some of them carry the spectral analysis not using remote sensors but using local hyperspectral spectroscopy only, thus not considering atmospheric variations that can happen in the acquisition procedure; 2) most of them rely on spectral signatures from pixels as features, which can be very noisy and error prone; 3) additionally, the classification of millions of pixels from a testing image can be computationally expensive; and 4) some methods, especially the one proposed by Lisita *et al.* [42], require semiautomatic classification, which can be too subjective and require an expert user to help classify samples. We discuss later in this manuscript how our proposed approach deals with such issues. However, to help understand the proposed approach, we discuss the dataset first in the next section.

IV. DATASET

The localization information of sensitive remote sensing data such as drug sites is not easy to acquire without the collaboration of law-enforcement organizations. This research is the result of a collaborative work with the Federal Police of Brazil, which provided us with annotations containing information from 800 places in a specific region in Brazil, called the *Marijuana Polygon*, where police operations against illegal crops took place in the past. The annotations contain the information of these areas from a period ranging from the years 2003 through 2012, such as date of the police operation, latitude, longitude, type of local (continent or island), city, area, number of plants, descriptions about the possible vegetation used to work as camouflage (i.e., to visually hide the illegal crops), among others.

After some data cleaning, we found the information to commercially acquire one high-resolution satellite image that covers the area of interest, composed by mosaicking four single images from different areas at the dates indicated in the annotations. The images cover 100 km², captured by an IKONOS sensor in September of 2008. They are composed by four multispectral bands (red, green, blue, and near-infrared) and a panchromatic band, with 0% cloud covering. The acquisition angle was 29.3°, and the spatial resolution was 3.2 m on each multispectral bands and 0.82 m in the panchromatic band. The images are in the GEOTIFF format with 11-b radiometric resolution, with UTM/WGS84 projection and orbital altitude of 684 km. The images are also accompanied by geometrical correction and RPC—rational polynomial coefficients—calculated from camera and sensor information.

Some of the four mosaicking images showed color variation in the scanning order, as the area of 100 km² was covered in four satellite sweeps, with the illumination due to the sun varying during such sweeps. Although the sun angle azimuth and sun angle elevation showed a maximum difference of 1 min between the acquisitions, we found different values of scan azimuth and nominal collection azimuth in the images. To minimize the differences of scan direction and acquisition azimuths, we applied the relative radiometric normalization [43], [44] method. For

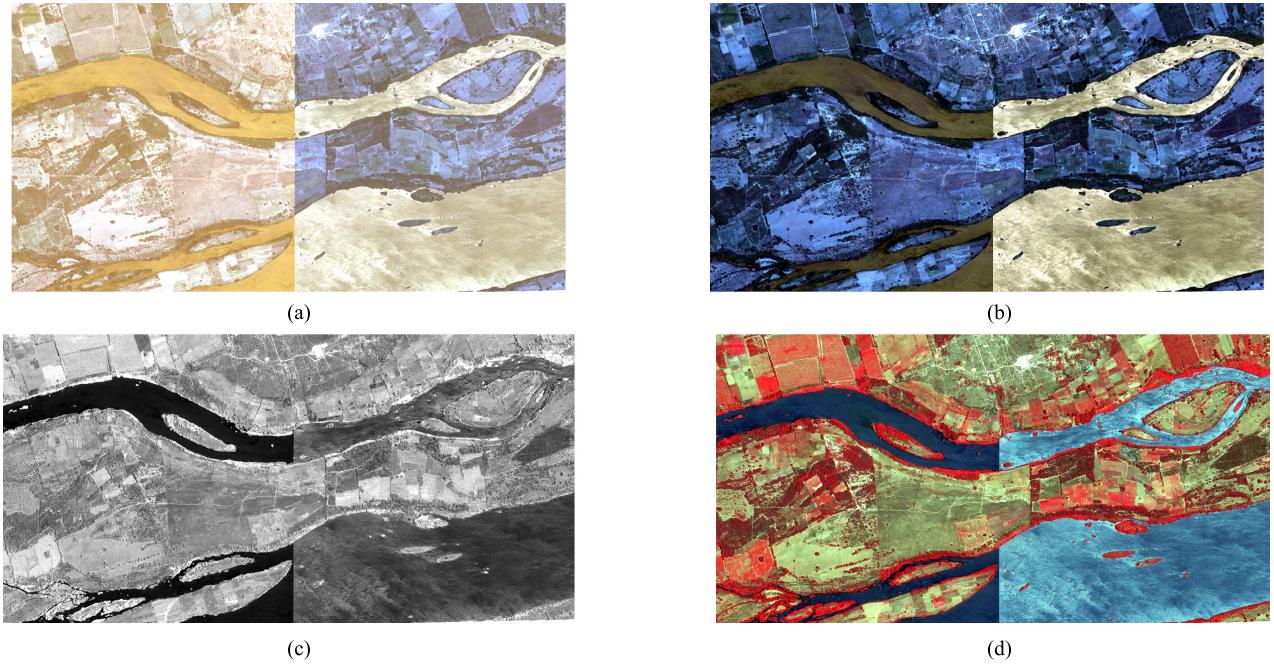


Fig. 1. Different versions of the image used to acquire the dataset. (a) Original image from the IKONOS satellite, (b) radiometric normalized, (c) radiometric normalized near-infrared channel, and (d) radiometric normalized false-color representation

that, we used the histogram-matching radiometric normalization approach, as it shows the smallest root mean square error for this sensor, as pointed out by Hong and Zhang [45].

The histogram-matching algorithm assumes that, with a large scene, the distribution of the intensity of Earth's radiation incident on each detector will be similar. Given two images, the reference and target images, we compute their histograms, and the method transforms the target image's histogram distribution into the specified histogram of the reference image, so that the radiometric appearance of the image to be transformed and the reference image become similar. Rakwatin *et al.* [46] describe the histogram-matching maps with the cumulative distribution function (CDF) of each detector to a reference CDF. A normalization lookup table is created for each detector to map every digital number (DN) x , for each sensor, onto the referenced DN x' . If $p_i(x)$ is the histogram of the output from the i th detector, the CDF of the i th detector $P_i(x)$ is

$$P_i(x) = \sum_{t=0}^x p_i(t). \quad (3)$$

According to Rakwatin *et al.* [46], the CDF is a nondecreasing function of x . The basic assumption is that the CDF of each detector is a monotonic function. For each output value x of the i th detector, the value x should satisfy the following:

$$P_r(x') = P_i(x') \quad (4)$$

where the subscript r refers to the reference detector. Therefore, a modified DN x' can be obtained as

$$x' = P_r^{-1}(P_i(x)). \quad (5)$$

To apply the histogram-matching radiometric normalization in our images, we chose, from the available images from the mosaic, the less corrupted one as our reference. Fig. 1 shows the original image used as the base to build the dataset and its variations used in the experiments. We use the near-infrared and false-color representations of images (see Fig. 1(c) and (d), respectively) to extract 30×30 resolution patches for the dataset. The majority of the crop sites are located on the islands located in the San Francisco river, which crosses the *Marijuana Polygon*. This happens for two reasons: 1) the dry weather of the Brazilian Caatinga, which makes it more difficult for the farmers to find water on land; and 2) by using islands as crop sites, the accessing and monitoring from security forces are harder and expensive. So, for patches' extraction, we considered only green areas and soil, removing patches containing river and buildings. We considered 126 areas containing *Cannabis Sativa* crops in advanced growth stage from the provided annotations and divided them into 63 regions on a set called bag_1 (patches from the islands on the left side of the river), and 63 regions on a set called bag_2 (patches from the islands on the right side of the river). As the area covered by bag_1 is smaller than that covered by bag_2 , when we created the subsets of the patches, we obtained 1149 and 9277 patches, respectively, creating a final dataset of 10426 images of size 30×30 , divided in the following three sets.

- 1) *Near_1*: These patches contain at least 75% of *Cannabis* crops pixels. We label these patches as positive samples. As the problem herein is a classification problem of very unbalanced classes, we found that 75% of marijuana pixels is a good parameter to choose the maximum number of relevant positive patches, in such a way to also avoid machine-learning classifier confusions as much as possible.

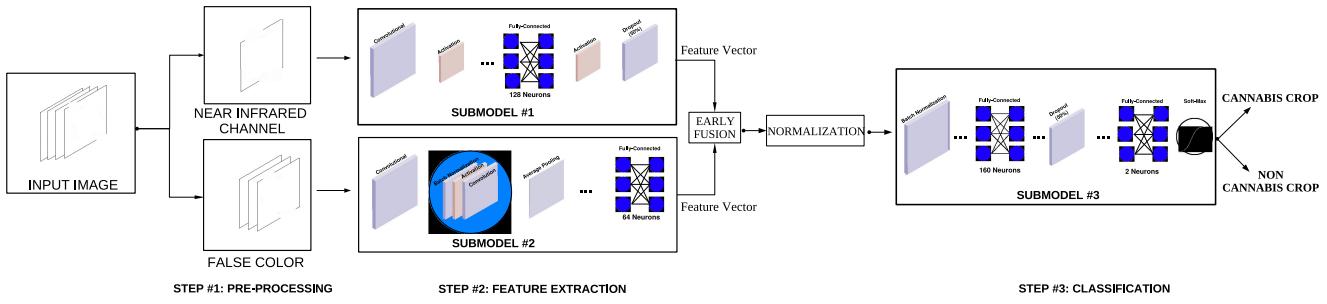


Fig. 2. Proposed data-driven fusion approach to detect the drug sites in remote sensing images. First, small patches from different input image-representations selected in a preprocessing step are fed to specific pretrained custom-tailored CNNs with different architectures. In the fully connected layer of these CNNs, feature vectors are extracted, concatenated, normalized, and further fed to another CNN for decision making.

- 2) *Near_2*: These patches contain green vegetation used as camouflage to hide positive samples and patches containing less than 75% of *Cannabis* crops pixels. We label these patches as negative samples.
- 3) *Near_3*: These patches contain the rest of Earth materials (except water and buildings), including the native vegetation of the Brazilian Caatinga. They are also considered negative samples.

Following this classification, we have bag_1 with 139 positive and 1010 negative samples, and bag_2 with 1028 positive samples and 8249 negative samples. In the next section, we discuss how to deal with this limited amount of data to train our proposed data-driven solution.

V. PROPOSED APPROACH

Our proposed approach herein, whose pipeline we show in Fig. 2, tackles the issues previously discussed in the related work section by focusing on the following aspects: 1) image acquisition by satellites; 2) patch-based analysis rather than pixel-wise analysis, in order to reduce the number of samples to be characterized and classified each time; 3) multiple and complementary input representations, which will highlight the areas of interest differently; 4) multiple data-driven characterization of the inputs, used to explore the discriminating characteristics of each representation in a complementary way; and 5) classification considering a normalized space, as the features come from different sources.

The proposed method works on image patches and combines the characterization output of different CNN architectures, applied over different image representations. These image representations were chosen in such a way that they highlight differently, but in a complementary way, the image's chlorophyll difference between the crop of interest and other Earth materials, including the native vegetation around the cropping area. The pipeline of the proposed method works in three steps: Step #1 is used as a preprocessing step, selecting and organizing channels in order to highlight green areas in a complementary way. Then, Step #2 will use specific pretrained deep-learning networks at the each of these representations, in order to learn complementary features that can better discriminate these crops. Finally, Step #3 uses another deep-learning network to classify these features. In the following sections, we show more details of the proposed

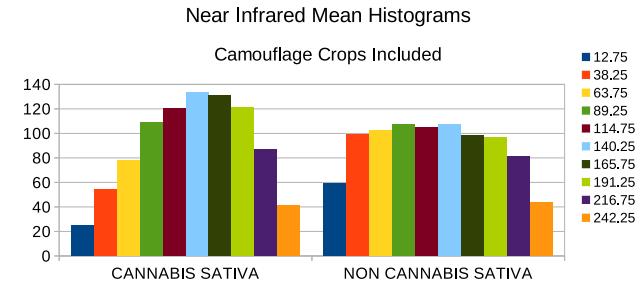


Fig. 3. Mean histogram of positive samples (i.e., *Cannabis Sativa* crops) and negative samples (i.e., non-*Cannabis Sativa* crops) in the near-infrared channel of the IKONOS 30×30 -pixel remote sensing image patches.

data-driven solution to the detection of *Cannabis Sativa* crops in remote sensing images.

A. Choice of Multiple Representations (Step #1)

To find out which inputs would be best for data-driven analysis of *Cannabis Sativa* crop's identification, we carried out an investigation in the dataset, as already described in Section IV, which is composed of four-channel IKONOS image patches. To probe possible discriminative features, we calculated the mean histogram of near-infrared channel in positive (i.e., *Cannabis Sativa* crops) and negative remaining areas. This helps us find out whether there is any spectral difference between *Cannabis Sativa* and non-*Cannabis Sativa* crops.

Fig. 3 shows that there is a discriminative feature in the near-infrared channel on higher bins to identify *Cannabis Sativa* crops in remote sensing images. This highlights the fact that, on the near-infrared band, the mean values of *Cannabis Sativa* crops are higher than those in other Earth materials. This can be explained as *Cannabis Sativa* leaves have a specific green emerald color and unusual shape, which was shown in previous work to have a peculiar spectral signature [37]–[39]. As the near-infrared channel is useful to discriminate different kinds of plants [47], the difference of *Cannabis Sativa* vegetation from other Earth materials tends to be detectable in this channel. However, as we are considering camouflage areas that contain less than 75% of the crops of interest (*Near_2* patches), there is some overlap in these higher bins. So, to implement a robust approach to tackle classification errors, in our proposed method, we chose to create an ensemble approach, focused on different representations that

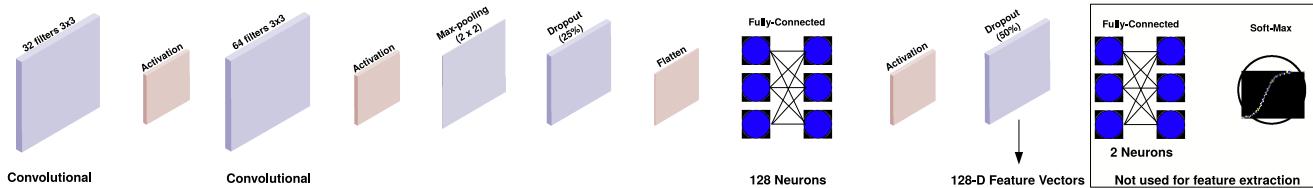


Fig. 4. Submodel #1 CNN architecture applied on 28×28 near-infrared image patches for *Cannabis Sativa* crop detection. We train this network from scratch and use it as a feature extractor to discriminate positive and negative samples using near-infrared image representations as input.

highlight differently green areas. These representations are then applied to different data-driven approaches that can work complementing each other, in order to avoid individual classification errors.

Upon these findings, we chose to focus on the following two image representations for our data-driven approach.

- 1) *Near-infrared*: This is the fourth multispectral band acquired by an IKONOS satellite. The light wavelength it captures is between 0.75 and $1.4\text{ }\mu\text{m}$.
- 2) *False color*: The false-color image representation is widely used to show images in color, acquired on visible and invisible parts of the electromagnetic spectrum. It is typically used to depict an object in colors that differs from a true-color image. In contrast to a normal true-color image, the false-color image sacrifices natural color disposition in order to ease the detection of features that are not readily discernible otherwise. The simplest way to create a false-color image is by choosing and reordering spectral bands, according to the physical properties of the object under investigation. We chose the false-color image representation using the following mapping: $BGRN \rightarrow NRG$, where N , R , and G are the near-infrared, red, and green image channels, respectively, with the blue band B being discarded. This representation is commonly called the *vegetation in red* false-color image and is widely used in remote sensing image processing [48].

B. Use of Multiple Network Architectures (Step #2)

With multiple input-image representations available, we devise two CNN architectures to learn complementary features, directly from these representations and from available training data (data-driven methods). In the next paragraphs, we discuss the adopted architectures—Submodels #1 and #2.

1) *Submodel #1*: This model was thought of to process near-infrared image representations and is built to process 28×28 image patches in its input layer. This input-image size is thought of to minimally represent the properties of our interest in a possible plantation area. The network's architecture comprises four main layers (two convolutional and two fully connected) and eight auxiliary layers. This network is trained from scratch to find out the discriminative features on training and validation data. After training the network, any 28×28 -pixel input can be represented with 128-dimensional features, as the data-driven description for near-infrared images. These features are extracted from the dropout layer located after the output

activation of the last but a fully connected layer of the trained network. Fig. 4 shows the disposition of layers in the proposed network.

We found the best parameters of the proposed Submodel #1 after some validation experiments, using 90% of the training data to train and 10% of the training data for validation, and they are as follows. Submodel #1 convolutional layer weights were initiated with the Xavier initialization method from Glorot and Bengio [49]; the learning rate was set to 10^{-2} and was kept constant. The optimization algorithm chosen was the ADAMAX of Kingma and Ba [50] with $\epsilon = 10^{-8}$. This network was trained using a batch size of 32 images through 20 epochs. The loss function chosen was the categorical cross entropy. Fig. 4 shows two dropout layers in the network: The first one eliminates neurons with 25% of probability and the last one with 50% of probability. For reference, there is no batch normalization in this network, and there is just one pooling layer, extracting the maximum value in a 2×2 window.

Submodel #1 network is somewhat inspired on the LENET [51] CNN for digit recognition, as it also deals with the same kind of input (grayscale 28×28 -pixel patches). However, there are some key differences in the proposed network structure, training steps, and final weights because of the following reasons: 1) our proposed Submodel #1 is trained with different data, resulting in different filter weights; 2) weights are initialized with Xavier initialization [49] instead of using a Gaussian distribution; 3) network hyperparameters are different from the original LENET, and the adopted optimization algorithm is the ADA-GRAD [52] rather than the steepest gradient descent; 4) we use dropout layers to avoid overfitting; and finally 5) the soft-max layer here identifies two classes of interest, rather than ten.

2) *Submodel #2*: This model was thought of to process false-color image representations, and its input layer accepts $32 \times 32 \times 3$ image patches. This is a very deep architecture with 41 main layers (39 convolutional and two fully connected), complemented by batch-normalization layers, activation layers, and concatenate layers whenever necessary, plus 3 pooling layers, arranged in an architecture such as the one shown in Fig. 5.

This architecture comprises a set of *dense blocks*, each of which contains several *composite functions*. The composite functions are formed by a batch-normalization layer, an activation (RELU) layer, and a convolutional layer, and the output of each composite function is concatenated with the outputs and inputs of the previous composite functions. The network comprises three dense blocks, with 13, 13, and 12 composition functions, respectively. The first dense block will process 32×32

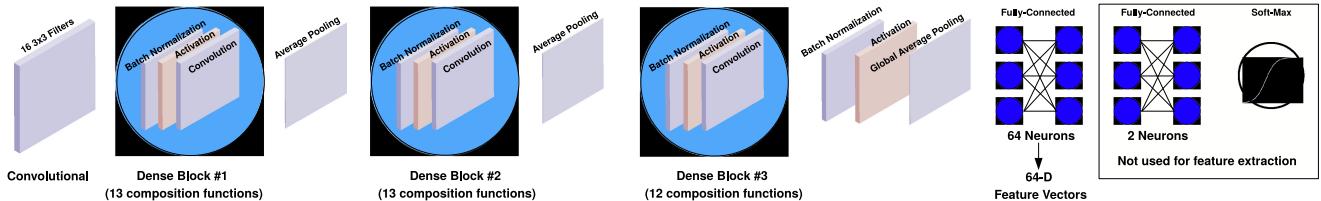


Fig. 5. Submodel #2 architecture applied on 32×32 false-color image patches for *Cannabis Sativa* crop detection. The architecture is mainly composed of several composition functions, formed by a batch-normalization layer, an activation (ReLU) layer, and a convolutional layer, with its output concatenated with the input and output of the previous composition functions. These composition functions are spread in three dense blocks, containing 13, 13, and 12 composition functions, respectively. Pooling layers connect the output of a dense block to the input of the next dense block. We train this network from scratch and use it as a feature extractor to discriminate positive and negative samples, using the false-color image representation as input.

feature maps. After the 13th composition function, an average pooling layer takes place and yields 16×16 feature maps, further processed by the second dense block. Another average pooling, after the 26th composition function, returns 8×8 feature maps. After 12 more composite functions from the third dense block, a batch normalization and activation take place. Then, a global average pooling will yield 488-dimensional vectors, which will flow to two fully connected layers for the final processing. The first one processes the data yielding 64-dimensional vectors. The last layer is typically used for classification along with a soft-max activation function, responsible for predicting output labels.

As with Submodel #1, this network is also trained from scratch to find out filter weights capable of discriminating *Cannabis Sativa* and non-*Cannabis Sativa* image patches. After training the network, any 32×32 false-color input patch can be represented by 64-dimensional features as the data-driven description for the false-color images. These features are extracted from the last but one fully connected layer of the trained network.

Submodel #2 convolutional weights were initiated with the initialization method of He *et al.* [53]; the initial learning rate was set to 10^{-2} and reduced by a factor of $\sqrt{0.1}$ when the validation loss stops improving after 10 epochs. The lower bound of the learning rate was fixed to 0.5×10^{-6} . The optimization algorithm chosen was the ADADELTA of Zeiler [54], with $\rho = 0.95$ and $\epsilon = 1e - 08$ as the regularization constant; the learning rate decay was defined as constant. Submodel #2 network was trained using a batch size of 64 images with the best validation loss defining the number of epochs after 300 total epochs. The chosen loss was the categorical cross entropy as with our proposed Submodel #1 CNN. As discussed earlier, there are three average pooling layers and 39 batch-normalization layers. There are no dropout layers in the network. The proposed network is somewhat inspired on the DENSENET [19] network, as the input for this network consists of 32×32 -pixel RGB patches, and feature maps from previous dense blocks are used in concatenation with the output of the following dense blocks. However, there are some key differences as follows on the network structure, training steps, and final weights: 1) the proposed Submodel #2 is trained with different data, resulting in different weights; 2) network hyperparameters are different from original DENSENET, and the adopted optimization algorithm is ADADELTA [54]; and 3) our fully connected layer extracts 64-dimensional vectors

from the last, but one RELU activation layer and our soft-max layer is set to identify two classes of interest.

Because the dataset used for this problem is naturally very unbalanced in the two classes and because our dataset has originally only 10 426 samples, we used the following procedure for data augmentation when training CNNs on these data, as CNNs normally require considerable amount of data for training. We perform a series of data-augmentation operations in the positive (i.e., *Cannabis Sativa*) data, with horizontal flips, vertical flips, rescales, zooms, and shear effects in order to give more positive data to the networks. Then, given a batch size BS , we randomly chose $\frac{BS}{2}$ positive patches and $\frac{BS}{2}$ negative patches from the augmented positive data and original negative data to create a batch of balanced samples to be processed by our CNNs. Doing so 10 000 times with $BS = 32$ in the initial training data, leads to 320 000 training images. We did the same 1000 times in the initial validation data to generate 32 000 validation images. Considering bag_1 and bag_2 images, our final dataset contains $(320\,000 + 32\,000) \times 2 = 704\,000$ images for *training* and *validating* Submodels #1 and #2. By using this procedure we may end up with some repeated images inside the training or inside the validation dataset, but they are never found in the same batch of images, and all the batches are balanced. It is needless to say that all the performed operations are for training the networks and do not mix *training*, *validation*, and *testing* data, whatsoever. To be suitable for the input layers of our proposed CNNs, we resized all the 30×30 images, using the Lanczos interpolation method [55], to be of the same input size of that of Submodels #1 and #2.

C. Classifying Fused Features (Step #3)

After training Submodels #1 and #2 on different image representations, our final model uses the generated features of both the approaches in an early fusion fashion [56], combining 128-dimensional feature vectors from Submodel #1 and 64-dimensional feature vectors from Submodel #2, generating $128 + 64 = 192$ -dimensional feature vectors for each input image. The final feature vector describes how complementary the two image representations are, according to the two different data-driven feature extractors.

As the final feature vector comes from different feature extractors applied on different image representations (i.e., different domain representations), the range of all the features should be

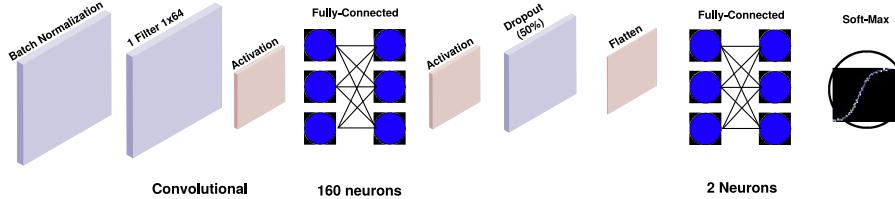


Fig. 6. Submodel #3 architecture focused on classifying normalized 192-dimensional vectors inputs from Submodels #1 and #2.

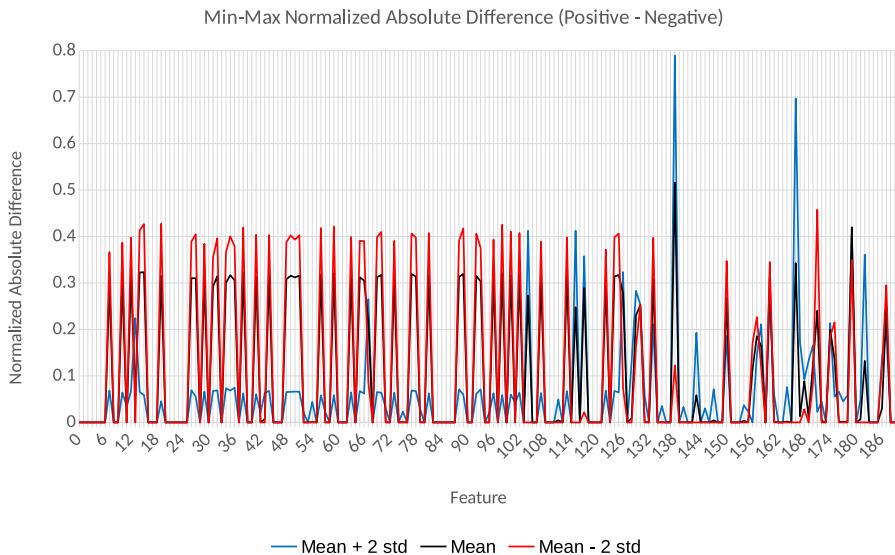


Fig. 7. Differences of the proposed 192-dimensional vectors for characterizing *Cannabis Sativa* and non-*Cannabis Sativa* samples. The differences illustrated in this figure show that the two classes have distinct values in some key positions.

normalized in such a way that each feature contributes approximately in a proportional way to the final feature vector. We chose to use the min–max normalization [57] method, which normalizes a vector in the following way:

$$f'_i = \frac{f_i - \min(f_i)}{\max(f_i) - \min(f_i)}, i \in \mathbb{N} \quad (6)$$

where f' is the final normalized vector, f the original vector, and $\min(f_i)$ and $\max(f_i)$ are the minimum and maximum values of vector f , respectively.

This normalization is an essential preprocessing step, which allows a classification algorithm to focus on structural similarities/dissimilarities of feature vectors, rather than on amplitude-driven ones. Fig. 7 shows the confidence-interval differences over the distribution of our 192 normalized neurons proposed CNN ensemble, considering the *Cannabis Sativa* and non-*Cannabis Sativa* crops. It is possible to realize from this figure that high peaks of differences are found in some features. They are present in the first 128 features of Submodel #1 and also in the remaining 64 from Submodel #2.

Finally, the decision making is carried out by another CNN, whose architecture is shown in Fig. 6. This new CNN, which we call Submodel #3, is composed of a batch-normalization layer, a convolutional layer with a filter of length 32 activated

by RELU, a fully connected layer with 160 neurons also activated by RELU, a 50% dropout operation, a flatten layer, and a fully connected layer followed by a softmax function, which will be used to classify the input. The CNN is trained using the RMSprop optimization algorithm [58] on a batch size of 32 images. The loss function chosen was the categorical cross entropy, and the initial learning rate was set to 0.01, which is reduced by a factor of $\sqrt{0.1}$ when the validation loss stops improving after 10 epochs. The lower bound of the learning rate was fixed to 0.5×10^{-6} , and the best validation loss will define the number of epochs after 100 total epochs.

Organizing the pipeline this way, we have a stronger classifier applied on the 192-dimensional vectors. This classifier will extract by convolutions the low-level details that discriminate the feature distributions of both the classes, rather than relying just on the distribution of vectors around a hyperplane as it happens with SVM classifiers. We show later in this paper that processing neuron activation differences from two CNNs by another CNN is effective in separating the classes of interest.

VI. EXPERIMENTAL SETUP

In this section, we present the experimental setup used to validate the proposed approach along with state-of-the-art feature engineering and data-driven image classification methods used

as baselines. We discuss the validation procedure with metrics and statistical tests, existing solutions (used to compare to ours), and implementation aspects of the proposed methods.

A. Metrics and Statistical Test

To compare the proposed algorithm with some baselines in the literature, we chose an experimental procedure called 2-fold cross validation, which is composed of two rounds: In the first one, one part of the dataset is used to train a classifier, and the other part is used for testing. Then, the parts are switched in the second round. We report mean metrics after these two rounds, using bag_1 and bag_2 images for this validation procedure.

We chose a set of metrics that take into account the unbalanced nature of this application. The first metric, the true positive rate (TPR), which is also known as *recall*, indicates the percentage of correctly classified positive (i.e., *Cannabis* crops) regions. We have

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7)$$

where TP (true positive) represents the number of patches correctly classified as *Cannabis Sativa* crops in the test phase and FN the number of positive samples wrongly labeled as negative.

In a real-world situation, we do not want to send security forces to wrong places. Thus, an important metric we want to minimize is the false positive rate (FPR), which indicates the percentage of incorrectly detected *Cannabis Sativa* crops. We have

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (8)$$

where FP (false positive) represents the number of negative patches wrongly classified as positive and TN the number of correctly labeled negative patches (i.e., non-*Cannabis Sativa* crops).

Another metric of interest is the normalized accuracy (NACC), which gives the quality of detection, based on TPR and TNR (true negative rate), with the last metric indicating the percentage of correctly located negative patches. We have

$$\text{NACC} = \frac{\text{TPR} + (1 - \text{FPR})}{2} \quad (9)$$

where $(1 - \text{FPR}) = \text{TNR}$.

In a complementary manner, we are also interested in calculating the precision, which is the fraction of the events in which the classifier correctly classified the *Cannabis Sativa* crops out of all the instances classified as such, accurately or wrongly. We have

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (10)$$

Finally, the most important metric, which will be used to rank the results, is the f-measure. This measure can be interpreted as the harmonic mean of precision and recall (i.e., TPR). It reaches its best value at 1, its worst score at 0, and it is calculated as follows:

$$f = 2 \times \frac{\text{Precision} \times \text{TPR}}{\text{Precision} + \text{TPR}}. \quad (11)$$

A series of statistical tests is performed to check whether the reported results are significantly different. We chose the McNemar [59] statistical test, which is a hypothesis test used to compare predictive accuracies of two classification models. It is applied on the outputs of the classification models, and it tests conditions to reject or accept the null hypothesis, which is assuming that the predicted class labels from the two algorithms have equal accuracy. We used the mid *p*-value variation of such test [60].

B. Baselines

In this paper, we carry out a very comprehensive comparative study on sensitive remote sensing classification by considering several image descriptors as baselines. We considered for this paper 17 different feature-engineering classical image representation approaches, with 6 being texture descriptors applied on the near-infrared image and 11 color descriptors applied on the false-color image representation. We also considered five data-driven state-of-the-art methods, with two of them trained from scratch and three of them fine-tuned on remote sensing data, totaling 22 baseline approaches.

As for the texture descriptors, we considered two descriptors from the work of Ferreira *et al.* [61], which are the variations of Gray-Level-Co-occurrence Matrices (GLCMs) (we call these approaches GLCM-MD and GLCM-MD-MS, respectively). Other GLCMs' variations considered in this paper are the original GLCM descriptor from Haralick *et al.* [62] (we call it GLCM-HARA) and its variation proposed by Mikkilineni *et al.* [63] (we call it GLCM-MIK). Finally, we considered the histogram of the oriented gradients from Dalal and Triggs [64], which we call HOG and the local configuration pattern (which we call LCP) of Zhao and Pietikainen [65].

All the color descriptors considered in this paper are applied on the false-color image representation and come from a comparative study performed by Bianconi *et al.* [66]. Because of the space requirements in the manuscript, we selected the following color descriptors: three versions of the three-marginal histograms (we call them TMH1, TMH2, and TMH3, respectively) from Pietikainen *et al.* [67], a two-marginal histogram (we call it 2MH) from Lepisto *et al.* [68], five versions of color statistics also called soft color descriptors (we call them CS1, CS2, CS3, CS4, and CS5, respectively) from Kukkonen *et al.* [69], Co-Occurrence Matrices and Color Percentiles (COM-CP) from Siliven *et al.* [70] and Integrative Co-Occurrence Matrices (ICOM) from Arvis *et al.* [71].

Finally, we compare our proposed approach with other five data-driven methods for classification. We trained two networks from scratch, using our remote sensing data of interest. The first is a CNN capable of classifying tiny image patches from the CIFAR dataset (we call it CIFAR), and the other is the RESNET from He *et al.* [17] (we call it RESNET). Then, we chose to fine-tune three networks on the remote sensing data of interest in this paper: the VGG-16 and VGG-19 networks, proposed by Simonyan and Zisserman [16] (we call them VGG-16 and VGG-19, respectively) and the INCEPTION version 3 network from Szegedy *et al.* [18] (we call it INCEPTION-V3).

TABLE I

VALIDATION EXPERIMENT, CONSIDERING *bag_1* IMAGES FOR TRAINING, *bag_2* IMAGES FOR TESTING AND VARYING IMAGE CHANNELS TO TRAIN AND TEST SUBMODEL#1 CNN

Rank	INPUT	<i>Metrics calculated considering PROPOSED_SUBMODEL1 after classifying bag_2 images</i>			
		F	NACC (%)	TPR (%)	FPR (%)
1	NEAR INFRARED CHANNEL	0.47	84.10	94.26	26.06
2	FALSE COLOR	0.34	74.52	92.90	43.86
3	GREEN CHANNEL	0.29	67.00	81.91	47.90
4	RED CHANNEL	0.24	60.47	96.89	75.15
5	BLUE CHANNEL	0.20	50.02	100.00	99.96

C. Implementation Aspects of the Proposed Methods

The implementation of Submodels #1, #2, and #3 followed the CNN architectures and characteristics described in Sections V-B and V-C. After training the CNNs of Submodels #1 and #2 by using the augmented data discussed previously in this paper, only the original training and testing data are used to extract feature vectors to train and test Submodel #3. For this classifier applied on the original (not augmented) data with very unbalanced nature, we used the SMOTE [72] approach in order to balance the number of samples in each class for CNN training.

We consider our approaches individually, classifying images with individual CNNs (we call these approaches PROPOSED_SUBMODEL1 and PROPOSED_SUBMODEL2), and fusing their descriptions using min–max normalization and classifying them with Submodel #3 (we call this approach PROPOSED_FUSION). To implement the proposed method, we used the KERAS library in Python [73], and we run our CNNs on an NVIDIA Tesla K80 GPU.

VII. EXPERIMENTAL RESULTS

In this section, we report results that validate our proposed approach and compare it with the existing baselines. We start by investigating the selection of the inputs of our proposed approach. Then, we show comparative experimental results of our approach against baseline methods, using hand-crafted and data-driven solutions. Finally, we include statistical tests to assess the contribution of our proposed approach in classifying drug sites in remote sensing images.

A. Inputs of the Proposed Approach

We start the experiments by showing an experiment result that helped us choose which input to use in each of our input submodels (Submodels #1 and #2). For that, we applied our submodels on different inputs in a validation experiment, considering *bag_1* images to train CNNs and *bag_2* images to test. Tables I and II shows the mean metrics of these experiments.

Table I shows that, for this problem of discriminating the green crops of interest, PROPOSED_SUBMODEL1 better works on the near-infrared channel. This happens because this network is designed to work on one-channel images only and because the near-infrared channel gives the best information possible about crops, natural vegetation, and other Earth materials. The promising results from the false-color representation can also be seen from this table, although this representation must be

TABLE II

VALIDATION EXPERIMENT, CONSIDERING *bag_1* IMAGES FOR TRAINING, *bag_2* IMAGES FOR TESTING AND VARYING IMAGE CHANNELS TO TRAIN AND TEST SUBMODEL#2 CNN

Rank	INPUT	<i>Metrics calculated considering PROPOSED_SUBMODEL2 after classifying bag_2 images</i>			
		F	NACC (%)	TPR (%)	FPR (%)
1	FALSE COLOR	0.58	88.55	93.48	16.39
2	NEAR INFRARED CHANNEL	0.46	82.38	89.40	24.65
3	BLUE CHANNEL	0.35	76.12	95.14	42.90
4	RED CHANNEL	0.22	57.01	99.51	85.50
5	GREEN CHANNEL	0.22	54.72	99.90	90.47

TABLE III

MEAN METRICS CALCULATED AFTER TWO ROUNDS OF EXPERIMENTS OF *Cannabis Sativa* CROP CLASSIFICATION AFTER THE 2-FOLD CROSS-VALIDATION PROCEDURE

Rank	Method	Information		<i>Metrics calculated after 2-fold Cross-Validation Experiments</i>			
		Type	Input	F	NACC (%)	TPR (%)	FPR (%)
1	PROPOSED_FUSION	Data-Driven	NI/FC	0.78 ± 0.19	92.16 ± 2.09	90.19 ± 3.42	5.87 ± 7.59
2	PROPOSED_SUBMODEL2	Data-Driven	FC	0.73 ± 0.21	90.32 ± 2.51	89.18 ± 6.07	8.54 ± 11.10
3	TMH3 [66]	Hand-crafted	FC	0.72 ± 0.21	86.93 ± 0.11	86.44 ± 11.34	8.56 ± 11.55
4	RESNET [17]	Data-Driven	FC	0.70 ± 0.18	87.37 ± 0.07	83.10 ± 10.69	8.34 ± 10.40
5	TMH1 [66]	Hand-crafted	FC	0.68 ± 0.28	87.67 ± 3.78	88.50 ± 10.20	13.15 ± 17.76
6	CS4 [68]	Hand-crafted	FC	0.68 ± 0.22	87.79 ± 1.00	87.36 ± 12.64	11.46 ± 14.64
7	PROPOSED_SUBMODEL1	Data-Driven	NI	0.63 ± 0.17	89.64 ± 5.15	94.10 ± 0.19	14.77 ± 10.11
8	CS3 [68]	Hand-crafted	FC	0.63 ± 0.18	86.65 ± 1.30	85.01 ± 10.33	12.90 ± 12.93
9	TMH2 [66]	Hand-crafted	FC	0.63 ± 0.29	83.57 ± 6.75	80.97 ± 4.62	13.81 ± 18.13
10	GLCM-MD-MS [60]	Hand-crafted	NI	0.62 ± 0.19	86.61 ± 3.59	86.71 ± 4.61	13.49 ± 11.79
11	2MH [67]	Hand-crafted	FC	0.60 ± 0.19	87.59 ± 5.31	92.19 ± 3.20	17.01 ± 13.83
12	CS1 [68]	Hand-crafted	FC	0.60 ± 0.10	88.81 ± 2.75	93.29 ± 0.67	15.66 ± 6.18
13	COM-CP [69]	Hand-crafted	FC	0.59 ± 0.24	87.35 ± 7.83	95.09 ± 3.22	20.39 ± 18.89
14	CIFAR	Data-Driven	FC	0.57 ± 0.13	87.16 ± 3.83	91.67 ± 0.13	17.34 ± 1.44
15	ICOM [70]	Hand-crafted	FC	0.57 ± 0.21	81.58 ± 0.58	80.47 ± 20.19	17.30 ± 21.38
16	GLCM-MD [60]	Hand-crafted	NI	0.57 ± 0.14	82.13 ± 3.40	78.26 ± 4.87	14.00 ± 11.68
17	GLCM-MIK [62]	Hand-crafted	NI	0.57 ± 0.19	85.52 ± 5.80	88.69 ± 3.33	18.20 ± 14.96
18	VGG-16 [16]	Data-Driven	FC	0.55 ± 0.09	78.74 ± 8.08	68.90 ± 31.31	11.41 ± 15.16
19	CS5 [68]	Hand-crafted	FC	0.53 ± 0.20	85.42 ± 8.12	94.62 ± 1.54	23.77 ± 17.79
20	VGG-19 [16]	Data-Driven	FC	0.51 ± 0.03	75.89 ± 9.63	62.56 ± 33.53	10.78 ± 14.27
21	LCP [64]	Hand-crafted	NI	0.46 ± 0.13	75.66 ± 3.64	70.93 ± 6.71	19.59 ± 13.99
22	HOG [63]	Hand-crafted	NI	0.44 ± 0.11	76.64 ± 2.75	79.01 ± 10.00	25.71 ± 15.50
23	GLCM-HARA [61]	Hand-crafted	NI	0.33 ± 0.02	67.00 ± 1.32	60.50 ± 11.28	26.50 ± 8.63
24	CS2 [68]	Hand-crafted	FC	0.32 ± 0.01	68.96 ± 5.64	76.84 ± 19.13	38.92 ± 7.84
	INCEPTION-V3 [18]	Data-Driven	FC	0.28 ± 0.04	64.77 ± 4.09	81.10 ± 7.87	51.55 ± 16.05

NI stands for near-infrared channel and FC for false-color image representation. The results are ordered by f-measure, and the best metrics are highlighted in gray.

converted to gray in order to fit the CNN input, losing this way important information. All the other individual channels result in significant performance losses.

Finally, the experimental results in Table II highlight that the false-color image representation showed to be the best image representation in highlighting the artifacts from illegal crops of interest in this paper. This happens because the characteristic green emerald color of *Cannabis Sativa* is better highlighted in this *NRG* color representation, which can make them discriminant, even if considering camouflage crops and natural vegetation.

B. Comparison With Baselines

We now discuss the comparative experimental results. Table III shows the experimental results of baseline the hand-crafted approaches, baseline data-driven descriptors, proposed data-driven fusion, and individual proposed CNN approaches.

From Table III, we can find that the best texture descriptor considering the mean f-measure and also the NACC is the multidirectional gray-level co-occurrence matrices statistical descriptor (GLCM-MD-MS) proposed in [61]. This descriptor performed the best in classifying near-infrared samples, as it considers more information using more neighboring directions and scales, which

will result in more matrices, containing richer texture information in which statistics are calculated. The best color descriptor is the three-marginal histograms version 3 (TMH3) from Pietikainen *et al.* [67]. This descriptor highlights the spectral differences between *Cannabis Sativa* and non-*Cannabis Sativa* crops, as shown in Fig. 3, by using histograms after the image is converted to the Ohta color space [74].

The ability of the CNNs to learn discriminative features from the data can be seen from the experimental results shown in Table III. The RESNET [17] is the best baseline CNN approach, as its architecture was thought of to feed more data to next layers, as it fuses information from the previous layers with the output of the current layers by summing up their feature maps. This way, the description is richer on false-color images, as more characterization information is passed to other layers. It is also worth discussing the poor results of fine-tuned networks on remote sensing data (VGG-16, VGG-19, and INCEPTION-V3). These networks were not effective in classifying these patches because the input layers of these networks require larger images, leading the upscaling of such images to lose too much semantical information. Also, the initial weights used by the network were thought of to classify another dataset (i.e., IMAGENET dataset).

The results from Table III show promising results for the proposed approach. The PROPOSED_SUBMODEL1 is the first to reach higher NACC than all baselines that classify near-infrared images. However, its false positive detection leaves it behind RESNET and other feature engineering approaches when considering the f-measure metric to assess the performance of such algorithms, as the precision decays with a high false positive rate.

The method PROPOSED_SUBMODEL2, a CNN applied on false-color images, showed the best NACC and the best f-measure if considering only individual approaches. This network describes well *Cannabis Sativa* and non-*Cannabis Sativa* crops, as its feed-forwarding learning involves concatenating feature maps from the previous layers, passing them to the following layers. This way, more and richer information about patches flow through the network, resulting in a more discriminative learning process.

Finally, the use of complementary information from multiple image representations (near-infrared and false-color image representations) in multiple CNN architectures (PROPOSED_SUBMODEL1 and PROPOSED_SUBMODEL2), as proposed in this paper, outperforms all the previous methods. The complementary information of these two model descriptions decreases the false positive rate of PROPOSED_SUBMODEL1 and PROPOSED_SUBMODEL2 at the same time. This results in a final model, which showed the highest f-measure for all the experiments (O.78), the highest NACC (92.16%), the lowest false positive rates and, consequently, the highest TNRs of all the experiments (FPR = 5.87% and TNR = 94.13%). This highlights the promising use of such proposed method to aid police efforts in finding and monitoring drug sites.

We also have compared the running time of CNN training, considering the proposed approaches and baselines CNNs. Such results can be found in Table IV below.

TABLE IV
RUNNING TIME FOR TRAINING SOME OF THE CNNS
CONSIDERED IN THIS PAPER

Rank	APPROACH	INPUT	RUNNING TIME (s)
1	PROPOSED_SUBMODEL3	NI/FC	31.75
2	PROPOSED_SUBMODEL1	NI	2480.05
3	RESNET [17]	FC	15274.54
4	PROPOSED_SUBMODEL2	FC	22226.63
5	PROPOSED_FUSION	NI/FC	22258.11
6	INCEPTION_V3 [18]	FC	44748.00
7	CIFAR	FC	52167.00

NI stands for near-infrared channel and FC for false-color image representation.

From the experiments results reported in Table IV, it can be seen that the fastest method for training is the submodel PROPOSED_SUBMODEL3. This happens, as it is a shallow network applied on vectors. The second fastest method to train is another shallow network, the PROPOSED_SUBMODEL1, now applied on matrices (near-infrared image channels). The slowest methods are the ones with higher number of layers and complexity, and trained from scratch. These are the cases of the CIFAR and INCEPTION-V3 CNNs.

Finally, as our approach PROPOSED_FUSION is composed of three CNNs, two of them being independent (PROPOSED_SUBMODEL1 and PROPOSED_SUBMODEL2) and one applied on the others output (PROPOSED_SUBMODEL3), the running time of training the proposed approach is the running time of training the slowest individual CNN (PROPOSED_SUBMODEL2) plus the running time of training the CNN used for classification (PROPOSED_SUBMODEL3). This results in a total running time of 22258.11 s. This is approximately the half of the training running time of the networks with higher complexity, such as the INCEPTION-V3 CNN, but with a superior classification performance, as can be seen from the results shown in Table III.

C. Statistical Tests

We finally show McNemar statistical tests results in order to assess the contributions of our proposed approach. For that, we selected the top five best approaches, considering the best f-measure results from different description methodologies to compose our top five best approaches. We consider the best texture descriptor (GLCM-MD-MS), the best color descriptor (TMH3), the best individual CNN trained from scratch (our PROPOSED_SUBMODEL2), the best CNN fine-tuned on remote sensing data (VGG-19), and the proposed approach, using early fusion by normalizing descriptions from PROPOSED_SUBMODEL1 and PROPOSED_SUBMODEL2 (PROPOSED_FUSION).

After selecting the approaches for the statistical test, we used the 2-fold cross-validation experiment's binary classification outputs to perform two McNemar tests. We show the statistical tests results from the first experiment, which considers bag_1 images for training the classifier and bag_2 images for testing the classifier, in Table V.

TABLE V
MCNEMAR STATISTICAL TEST RESULTS CONSIDERING THE CLASSIFICATION
OUTPUTS OF THE APPROACHES IN THE FIRST EXPERIMENT
FROM THE 2-FOLD CROSS VALIDATION USED

Rank	Method	PROPOSED_FUSION	PROPOSED_SUBMODEL2	TMH3 [66]	GLCM-MD [60]	VGG-19 [16]	TOTAL
1	PROPOSED_FUSION	0	1	1	1	1	4
2	PROPOSED_SUBMODEL2	-1	0	1	1	1	2
3	TMH3 [66]	-1	-1	0	1	1	0
4	GLCM-MD [60]	-1	-1	-1	0	1	-2
5	VGG-19 [16]	-1	-1	-1	-1	0	-4

1 = Line method is better than column method
0 = Line method is equivalent to column method
-1 = Line method is worse than column method

TABLE VI
MCNEMAR STATISTICAL TEST RESULTS CONSIDERING THE CLASSIFICATION
OUTPUTS OF THE APPROACHES IN THE SECOND EXPERIMENT
FROM THE 2-FOLD CROSS VALIDATION USED

Rank	Method	PROPOSED_FUSION	PROPOSED_SUBMODEL2	TMH3 [66]	GLCM-MD [60]	VGG-19 [16]	TOTAL
1	PROPOSED_FUSION	0	1	1	1	1	4
2	PROPOSED_SUBMODEL2	-1	0	1	1	1	2
3	TMH3 [66]	-1	-1	0	1	1	0
4	GLCM-MD [60]	-1	-1	-1	0	0	-3
5	VGG-19 [16]	-1	-1	-1	0	0	-3

1 = Line method is better than column method
0 = Line method is equivalent to column method
-1 = Line method is worse than column method

The results in Table V show that our proposed fusion approach (PROPOSED_FUSION) has statistical significant difference when compared to all the other approaches considered, including our proposed individual model on false-color images, called PROPOSED_SUBMODEL2. Finally, Table VI shows the statistical test results of the second experiment, which uses bag_2 images for training the classifier and bag_1 images for testing the classifier. For this case, the null hypothesis is also rejected when comparing the performance of our proposed approach against all other approaches considered, including the second best approach (PROPOSED_SUBMODEL2).

VIII. CONCLUSION AND FUTURE WORK

Classifying sensitive remote sensing data, such as drug crop sites, is a challenging task because of several reasons such as: 1) the difficulty of acquiring proper annotated datasets; 2) coping with acquisition errors; 3) the spectral response similarity of drug crops and natural green camouflage; and 4) using spectral response of the remote sensed material only (i.e., pixel values),

as proposed by most solutions in prior art; this also makes data collection to train classifiers a difficult task.

In this paper, we dealt with some of these problems and performed what can be considered one of the most comprehensive studies in image processing of such sensitive remote sensing data. By doing an extensive study on a Brazilian Federal Police *Cannabis* drug apprehension set of records, we built what is, as far as we know, the first public dataset of remote sensing *Cannabis* and non-*Cannabis* crops images. From the dataset acquisition, preparation, characterization, and classification, we introduced novel contributions to this field, such as locating the areas of interest, fixing acquisition errors, finding the best representations to discriminate the crop of interest, among others.

Especially, our proposed data-driven approach to characterize these drug crops takes into account the complementary information from the near-infrared channel and false-color image representations. These two representations, combined to better discriminate the vegetation types, are used as input to an ensemble of classifiers containing two specific CNNs, capable of discriminating the chlorophyll information of these crops from the rest of Earth materials. Although the near-infrared data is part of the false color, we found CNNs with different architectures trained on these two kinds of data, which can work in a complementary way in a fusion (ensemble) scheme, which can be robust against failures of individual classifiers. This ensemble of CNNs using multiple architectures on multiple image representations works as a powerful feature extractor, later combined with an SVM classifier for decision making.

Our proposed method showed the highest mean f-measure, the best accuracy, and one of the lowest false detections, being a promising path toward the development of machine-learning approaches focused on drug crops' detection in remote sensing images. Although the proposed ensemble had a small mean improvement than that of the best individual classifier, it is designed in such a way to correct individual classifier confusions. So, it can be robust in real-world situations, where millions of patches are used for classification and acquisition noise can happen. For the specific case of detecting *Cannabis Sativa* crops, this research could also enable the collection of evidence from satellite imagery even after a plantation ceases to exist, by means of looking back at imagery history collected by different satellites for specific places. This would open a whole new avenue for prosecution and evidence gathering, as currently all we can do is rely on police investigations and apprehensions on site.

Notwithstanding, the research in this front still has important challenges to deal with in possible future work. First of all, other automatic/semitautomatic image-selection methods could be applied to select the most discriminative patches inputs to be used with tailored CNNs. Thus, segmentation or even different patch-selection methods must be thought of to find out the best way to include only positive or negative classes in the analyzed patches. This is of special interest in this study, as some false positives detected by our approach contain some *Cannabis Sativa* traces; i.e., some $Near_2$ patches containing camouflage and *Cannabis Sativa* samples in the same patch were classified as containing drug crops, given our chosen requirement for truly tagging a positive sample. Another interesting future work consists of

studying the effect of other image representations as input to the CNNs, as vegetation indices [75] have promising complementary information that could be used in a stacked fashion, with other image representations in a single network, or even used in an ensemble fashion with their own customized CNNs, as proposed in this paper. Finally, other promising future work is designing *ad hoc* CNNs for different representations of 30×30 blocks.

ACKNOWLEDGMENT

The authors thank the Brazilian Federal Police for the collaboration and data sharing without which this research would not have been possible.

REFERENCES

- [1] M. Asbridge, J. A. Hayden, and J. L. Cartwright, "Acute cannabis consumption and motor vehicle collision risk: Systematic review of observational studies and meta-analysis," *Brit. Med. J.*, vol. 344, 2012, Art. no. e536.
- [2] S. K. Nayak *et al.*, "Effect of cannabis consumption on ANS and conduction pathway of heart of Indian paddy field workers," in *Proc. IEEE Annu. India Conf.*, 2016, pp. 1–5.
- [3] G. Hulse, E. Milne, D. English, and C. Holman, "The relationship between maternal use of heroin and methadone and infant birth weight," *Addiction*, vol. 92, no. 11, pp. 1571–1579, 1997.
- [4] United Nations Office on Drugs and Crime, "World drug report 2017," United Nations, New York, NY, USA, Tech. Rep., May 2017. [Online]. Available: https://www.unodc.org/wdr2017/field/Booklet_3_Plantbased.pdf
- [5] United Nations Office on Drugs and Crime, "World drug report 2009," United Nations New York, NY, USA, 2009. [Online]. Available: http://www.unodc.org/documents/wdr/WDR_2009/WDR2009_eng_web.pdf
- [6] G. Ramsey and J. Walsh, "Uruguay is creating the world's first nationwide regulated cannabis market," 2017. [Online]. Available: <http://www.druglawreform.info/en/weblog/item/7594-uruguay-is-creating-the-world-s-first-nationwide-regulated-cannabis-market>
- [7] A. P. B. Lopes, S. E. F. D. Avila, A. N. A. Peixoto, R. S. Oliveira, M. D. M. Coelho, and A. D. A. Araújo, "Nude detection in video using bag-of-visual-features," in *Proc. Brazilian Symp. Comput. Graph. Image Process.*, Oct. 2009, pp. 224–231.
- [8] S. Ozkan, E. Esen, I. Atil, and G. B. Akar, "Scene nudity level detection with deep nets," in *Proc. Signal Process. Commun. Appl. Conf. (SIU)*, May 2016, pp. 2069–2072.
- [9] E. Yiallourou, R. Demetriou, and A. Lanitis, "On the detection of images containing child-pornographic material," in *Proc. Int. Conf. Telecommun.*, May 2017, pp. 1–5.
- [10] A. Ulges and A. Stahl, "Automatic detection of child pornography using color visual words," in *Proc. Int. Conf. Multimedia Expo.*, Jul. 2011, pp. 1–6.
- [11] H. Lang and H. Ling, "Classifying covert photographs," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2012, pp. 1178–1185.
- [12] H. Lang and H. Ling, "Covert photo classification by fusing image features and visual attributes," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 2996–3008, Oct. 2015.
- [13] N. Xie, H. Ling, and W. Hu, "Image set classification using multi-layer multiple instance learning with application to cannabis website classification," in *Proc. Int. Conf. Web Intell. Intell. Agent Technol.*, Aug 2010, vol. 1, pp. 58–65.
- [14] N. Xie, X. Li, X. Zhang, W. Hu, and J. Z. Wang, "Boosted cannabis image recognition," in *Proc. Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
- [15] Y. Wang, N. Xie, W. Hu, and J. Yang, "Multi-modal multiple-instance learning with the application to the cannabis webpage recognition," in *Proc. Asian Conf. Pattern Recognit.*, Nov. 2011, pp. 105–109.
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [18] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. Int. Conf. Comput. Vision Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [19] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jul. 2017, pp. 2261–2269.
- [20] H. Santoso, H. Tani, and X. Wang, "A simple method for detection and counting of oil palm trees using high-resolution multispectral satellite imagery," *Int. J. Remote Sens.*, vol. 37, no. 21, pp. 5122–5134, 2016.
- [21] G. Camps-Valls *et al.*, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, Jul. 2004.
- [22] R. Khatami, G. Mountrekis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sens. Environ.*, vol. 177, pp. 89–100, 2016.
- [23] J. Munoz-Mari *et al.*, "Hyperlabelme : A web platform for benchmarking remote-sensing image classifiers," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 79–85, Dec. 2017.
- [24] J. P. Guerschman, J. M. Paruelo, C. D. Bella, M. C. Giallorenzi, and F. Pacin, "Land cover classification in the Argentine Pampas using multi-temporal Landsat TM data," *Int. J. Remote Sens.*, vol. 24, no. 17, pp. 3381–3402, 2003.
- [25] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "Analysis of time-series spectral index data to enhance crop identification over a Mediterranean rural landscape," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1508–1512, Sep. 2017.
- [26] Q. An, W. Gao, B. Yang, J. Wu, L. Yu, and Z. Liu, "Research on feature selection method oriented to crop identification using remote sensing image classification," in *Proc. Int. Conf. Fuzzy Syst. Knowl. Discovery*, Aug. 2009, vol. 5, pp. 426–432.
- [27] S. Mulyono, M. I. Fanany, and T. Basaruddin, "Genetic algorithm based new sequence of principal component regression (GA-NSPCR) for feature selection and yield prediction using hyperspectral remote sensing data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 4198–4201.
- [28] Z. Zhang, E. Pasolli, H. L. Yang, and M. M. Crawford, "Multimetric active learning for classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 1007–1011, Jul. 2016.
- [29] Y. Zhang, H. L. Yang, S. Prasad, E. Pasolli, J. Jung, and M. Crawford, "Ensemble multiple kernel active learning for classification of multisource remote sensing data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 845–858, Feb. 2015.
- [30] D. Ming, T. Zhou, M. Wang, and T. Tan, "Land cover classification using random forest with genetic algorithm-based parameter optimization," *J. Appl. Remote Sens.*, vol. 10, pp. 1–16, 2016.
- [31] P. Du, J. Xia, J. Chanussot, and X. He, "Hyperspectral remote sensing image classification based on the integration of support vector machine and random forest," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 174–177.
- [32] A. Bailly, D. Arvor, L. Chapel, and R. Tavenard, "Classification of MODIS time series with dense bag-of-temporal-sift-words: Application to crop-land mapping in the Brazilian Amazon," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2016, pp. 2300–2303.
- [33] Q. Zhu, Y. Zhong, B. Zhao, G. S. Xia, and L. Zhang, "Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, Jun. 2016.
- [34] S. W. Chen and C. S. Tao, "Multi-temporal PolSAR crops classification using polarimetric-feature-driven deep convolutional neural network," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, May 2017, pp. 1–4.
- [35] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.
- [36] C. Giri, Ed., *Remote Sensing of Land Use and Land Cover*. Boca Raton, FL, USA: CRC Press, 2012.
- [37] C. Daughtry and C. Walthall, "Spectral discrimination of Cannabis sativa L. leaves and canopies," *Remote Sens. Environ.*, vol. 64, no. 2, pp. 192–201, 1998.
- [38] C. Walthall *et al.*, "What do we know about the spectral signatures of illegal cannabis cultivation?" in *Proc. Office Nat. Drug Control Policy*, 2003, pp. 1–15.
- [39] C. Walthall *et al.*, "Detection of illegal cannabis cultivation using remote sensing," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, 2006, pp. 2281–2284.

- [40] B. Thiessen, "Satellite detection of cannabis sativa- outdoor grow operations," Can. Police Res. Center, Ottawa, Ontario, Canada, Tech. Rep. TR-04-2007, Mar. 2007.
- [41] I. Azaria, N. Goldshleger, E. Ben-Dor, and R. Bar-Hamburger, "Detection of cannabis plants by hyper-spectral remote sensing means," in *Proc. EARSeL Imag. Spectroscopy SIG Workshop*, 2009, pp. 1–6.
- [42] A. Lisita, E. Sano, and L. Durieux, "Identifying potential areas of cannabis sativa plantations using object-based image analysis of spot-5 satellite data," *Int. J. Remote Sens.*, vol. 34, no. 15, pp. 5409–5428, 2013.
- [43] X. Yang and C. Lo, "Relative radiometric normalization performance for change detection from multi-date satellite images," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 8, pp. 967–980, 2000.
- [44] Z. Yang and R. Mueller, "Unbiased histogram matching quality measure for optimal radiometric normalization," in *Proc. Annu. Conf. Amer. Soc. Photogramm. Remote Sens.*, 2008, pp. 1–12.
- [45] G. Hong and Y. Zhang, "Radiometric normalization of IKONOS image using QuickBird image for urban area change detection," in *Proc. Int. Symp. Remote Sens. Data Fusion Urban Areas*, 2005, pp. 1–5.
- [46] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Stripe noise reduction in MODIS data by combining histogram matching with facet filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1844–1856, Jun. 2007.
- [47] J. R. Eastman, *Guide to GIS and Image Processing*. Clark University, Worcester, MA, USA, vol. 2, 2001.
- [48] Centre for Remote Imaging, Sensing and Processing (CRISP), "Interpretation of optical images," 2001. [Online]. Available: https://crisp.nus.edu.sg/research/tutorial/opt_int.htm
- [49] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist. (series Proceedings of Machine Learning Research)*, Y. W. Teh and M. Titterington, Eds., Chia Laguna Resort, Sardinia, Italy, 2010, vol. 9, pp. 249–256.
- [50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–15.
- [51] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [52] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Jul. 2011.
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1026–1034.
- [54] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: <http://arxiv.org/abs/1212.5701>
- [55] K. Turkowski, *Graphics Gems*. San Diego, CA, USA: Academic, 1990.
- [56] C. G. M. Snoek, "Early versus late fusion in semantic video analysis," in *Proc. Int. Conf. Multimedia*, Nov. 2005, pp. 399–402.
- [57] S. G. K. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *Int. Adv. Res. J. Sci., Eng. Technol.*, pp. 20–22, Mar. 2015.
- [58] T. Tieleman and G. Hinton, "Lecture 6.5-RMSProp: Divide the gradient by a running average of its recent magnitude," COURSERA: Neural Networks for Machine Learning, 2012.
- [59] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [60] M. W. Fagerland, S. Lydersen, and P. Laake, "The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional," *BMC Med. Res. Methodol.*, vol. 13, no. 1, pp. 1–8, Jul. 2013.
- [61] A. Ferreira, L. C. Navarro, G. Pinheiro, J. A. dos Santos, and A. Rocha, "Laser printer attribution: Exploring new features and beyond," *Forensic Sci. Int.*, vol. 247, pp. 105–125, 2015.
- [62] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [63] A. K. Mikkilineni, O. Arslan, P.-J. Chiang, R. M. Kumontoy, J. P. Allebach, and G. T.-C. Chiu, "Printer forensics using SVM techniques," in *Proc. Int. Conf. Digital Printing Technol.*, Oct. 2005, pp. 223–226.
- [64] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. Int. Conf. Comput. Vision Pattern Recognit.*, Jun. 2005, pp. 886–893.
- [65] Y. Guo, G. Zhao, and M. Pietikäinen, "Texture classification using a linear configuration model based descriptor," in *Proc. Brit. Mach. Vision Conf.*, 2011, pp. 119.1–119.10.
- [66] F. Bianconi, R. W. Harvey, P. Southam, and A. Fernandez, "Theoretical and experimental comparison of different approaches for color texture classification," *J. Electron. Imag.*, vol. 20, pp. 1–20, 2011.
- [67] M. Pietikäinen, S. Nieminen, E. Marszałek, and T. Ojala, "Accurate color discrimination with classification based on feature distributions," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 1996, vol. 3, pp. 833–838.
- [68] L. Lepistö, I. Kunttu, and A. Visa, "Color-based classification of natural rock images using classifier combinations," in *Scand. Conf. Image Anal.*, H. Kalviainen, J. Parkkinen, and A. Kaarna, Eds., 2005, pp. 901–909.
- [69] J. P. S. P. Saku Kukkonen and H. A. Kaelvæinen, "Color features for quality control in ceramic tile industry," *Opt. Eng.*, vol. 40, pp. 40-1–40-8, 2001.
- [70] O. Silvén, M. Niskanen, and H. Kauppinen, "Wood inspection with non-supervised clustering," *Mach. Vision Appl.*, vol. 13, no. 5, pp. 275–285, Mar. 2003.
- [71] V. Arvis, C. Debain, M. Berducat, and A. Benassi, "Generalization of the cooccurrence matrix for colour images: Application to colour texture classification," *Image Anal. Stereol.*, vol. 23, no. 1, pp. 63–72, 2011.
- [72] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
- [73] F. Chollet *et al.*, "Keras," 2015. [Online]. Available: <https://github.com/fchollet/keras>
- [74] Y.-I. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Comput. Graph. Image Process.*, vol. 13, no. 3, pp. 222–241, 1980.
- [75] A. Bannari, D. Morin, F. Bonn, and A. Huete, "A review of vegetation indices," *Remote Sens. Rev.*, vol. 13, no. 1-2, pp. 95–120, 1995.