

Article

Hyperspectral Bare Soil Index (HBSI): Mapping Soil Using an Ensemble of Spectral Indices in Machine Learning Environment

Eric Ariel L. Salas *  and Sakthi Subburayalu Kumaran 

Agricultural Research Development Program (ARDP), Central State University, Wilberforce, OH 45384, USA; ssubburayalu@centralstate.edu

* Correspondence: esalas@centralstate.edu

Abstract: Spectral remote-sensing indices based on visible, NIR, and SWIR wavelengths are useful in predicting spatial patterns of bare soil. However, identifying an effective combination of informative wavelengths or spectral indices for mapping bare soil in a complex urban/agricultural region is still a challenge. In this study, we developed a new bare-soil index, the Hyperspectral Bare Soil Index (HBSI), to improve the accuracy of bare-soil remote-sensing mapping. We tested the HBSI using the high-spectral-resolution AVIRIS-NG and Sentinel-2 multispectral images. We applied an ensemble modeling approach, consisting of random forest (RF) and support vector machine (SVM), to classify bare soil. We found that the HBSI outperformed other existing bare-soil indices with over 91% accuracy for Sentinel-2 and AVIRIS-NG. Furthermore, the combination of the HBSI and the normalized difference vegetation index (NDVI) showed a better performance in bare-soil classification, with >92% accuracy for Sentinel-2 and >97% accuracy for AVIRIS-NG images. Also, the RF-SVM ensemble surpassed the performance of the individual models. The novelty of HBSI is due to its development, since it utilizes the blue band in addition to the NIR and SWIR2 bands from the high-spectral-resolution data from AVIRIS-NG to improve the accuracy of bare-soil mapping.

Keywords: bare-soil index; hyperspectral bare soil index; soil mapping; urban–agricultural complex



Citation: Salas, E.A.L.; Kumaran, S.S. Hyperspectral Bare Soil Index (HBSI): Mapping Soil Using an Ensemble of Spectral Indices in Machine Learning Environment. *Land* **2023**, *12*, 1375. <https://doi.org/10.3390/land12071375>

Academic Editors: Zamir Libohova, Kabindra Adhikari, Michele Duarte De Menezes and Subramanian Dharumarajan

Received: 13 June 2023
Revised: 6 July 2023
Accepted: 8 July 2023
Published: 10 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bare soil is crucial for an understanding of the ecosystem structure, and has become a key driver of ecological functioning [1]. The dynamics of bare soil is a focus in soil evaporation studies and in the quantification of the overall water balance [2,3], the prediction of dust deposition [4], and the assessment of urban development [5]. Bare-soil features, such as albedo, are widely used in climate models for the retrieval of satellite-based land surface properties [6]. Given the importance of bare soil, there is a need to enhance current predictive methods and use publicly available spatial data to produce bare-soil dynamic mapping estimations.

Remote sensing (RS) data have long been documented as an important variable in predictive soil mapping [7]. For this purpose, soil indices (SI) derived from RS images have been consistently utilized [7]. However, misclassification of bare soil often occurs due to the challenges in accurately capturing the diverse spectral characteristics of bare-soil parcels using these soil indices. SIs are specifically designed to detect bare-soil parcels using various wavelength bands that are believed to be sensitive to bare soil, including visible (VIS) to near-infrared (NIR: 750–850 nm) and shortwave infrared (SWIR: 900–2500 nm) [8,9]. The use of VIS-NIR-SWIR for the spectral analysis of soil characteristics (e.g., soil organic carbon, pH, bulk texture) has grown in popularity over the past decades [10]. The bare-soil index (BSI), for example, was formed using the SWIR2 (~2080–2350 nm) and green band (500–600 nm), and is used to map potential areas of soil degradation and erosion [11]. BSI, in combination with the normalized difference vegetation index (NDVI), also discriminates exposed soil from vegetation cover when applied to an intensive agricultural area [12].

The bareness index (BI), on the other hand, utilizes the red (600–700 nm), NIR, and SWIR1 (~1570–1650 nm) to detect non-vegetated and urban area classes [13]. Other soil indices only used a combination of red and NIR regions [14] or red and thermal infrared (TIR: 10,000 to 11,200 nm) [15] for the detection of barren soil. Overall, these SIs have been widely utilized in land use-landcover (LULC) studies, specifically to distinguish bare land features from urban classes. However, in LULC classification, bare soil is frequently misclassified as an urban area, and vice versa, because of their overlapping spectral characteristics, and this can lead to challenges in accurate differentiation [16]. This misclassification issue has significantly hindered the effective mapping of urban areas using remotely sensed data, underscoring the need for extensive research and innovative approaches. To address this challenge, it is important to develop algorithms that can effectively identify and highlight the spectral bands that are proven to be the most useful for detecting bare soil. These algorithms must be capable of differentiating between bare soil and urban areas even in complex urban/agricultural regions, where the spectral signatures may vary due to diverse landcover compositions and environmental conditions.

Several studies have been published that used VIS-NIR and SWIR for soil characterization [17]. These wavelengths are known to provide valuable information about the reflectance properties of soil, allowing researchers to extract meaningful insights related to soil composition, moisture content, and other key parameters [17,18]. However, few attempts have been made to identify effective combinations of spectral bands and indices for mapping bare soil in a complex urban/agricultural region. By exploring novel combinations of spectral bands and indices, and developing advanced analytical techniques, we could enhance the precision and reliability of bare-soil mapping in urban/agricultural regions. Therefore, the major goal of this study was to identify and map bare-soil patches in an urban/agricultural site located in Anand, India using a new bare-soil index. To that aim, we tested the new soil index, compared it with existing soil indices, and used machine learning algorithms to classify two types of images: multispectral and hyperspectral images. We used publicly accessible high-spectral-resolution AVIRIS-NG and multispectral Sentinel-2 multispectral images as a low-cost and efficient method of mapping bare soil. We put a special emphasis on the dependability of the new bare-soil index that we developed, as well as on the combination of spectral indices for successful bare-soil identification and mapping.

2. Materials and Methods

2.1. Study Area

We selected the Anand District in the state of Gujarat, India as the study area (Figure 1). The site covers approximately 216 square kilometers (21,600 ha). The district falls within the Western Plain and Hill agroecological sub-region (ICAR). The average altitude of this region is 43 m above mean sea-level (MSL). The soil types in this region fall into the following two main categories: sandy loam and clay loam soils. The normal rainfall in this region is 687 mm, generated by the southwest monsoon (June–September). The major crops grown in the district include kharif-season irrigated rice, rabi-season irrigated and rainfed wheat, and kharif-season irrigated pearl millet, tobacco, cotton, and vegetables (potato, brinjal, tomato, and cabbage). The investigated site represents farmlands with diverse agricultural management and land-use systems.

2.2. Remote Sensing Dataset

We used available hyperspectral airborne visible infrared-imaging spectrometer-next-generation (AVIRIS-NG) imagery with a nominal ground resolution of 4 m to map the bare soil, taken on 26 March 2018. AVIRIS-NG samples 430 contiguous bands between 380 nm and 2510 nm at approximately 5 nm spectral resolution. An ortho-corrected and atmospherically corrected reflectance dataset (L2) for the study area is archived through NASA (<https://aviris-ng.jpl.nasa.gov>, accessed on 20 November 2022). We also used Sentinel-2 multispectral images available from the European Space Agency (ESA) Sentinels Scientific

Data Hub (<https://scihub.copernicus.eu/>, accessed on 20 November 2022). Sentinel-2 has 13 spectral bands: 4 bands at 10 m resolution, 6 bands at 20 m resolution and 3 bands at 60 m spatial resolution. The orbital swath width is 290 km. In this work, we downloaded the associated Sentinel-2 level 2A scenes taken on 20 March 2018 that were cloud-free and available for all sampling locations. Distributed level 2 products were atmospherically corrected by the Sen2Cor package (<https://step.esa.int/main/snap-supported-plugins/sen2cor/>, accessed on 20 November 2022). Both images matched up with, and were within, the field campaign days. We resampled all bands that were in 20 m resolution to 10 m to be consistent with the four native bands (band 2 in blue, band 3 in green, band 4 in red, and band 8 in NIR).



Figure 1. (a) Map showing the location of the study site in the Anand District in Gujarat, India. Examples of farmlands with sparse vegetation and bare-soil conditions are shown in (b,c), respectively.

2.3. Image Indices

We utilized RS image indices to differentiate land features (Table 1). Six vegetation indices (VI) including the normalized difference vegetation index (NDVI) [19], modified soil adjusted vegetation index 2 (MSAVI2) [20], soil adjusted vegetation index (SAVI) [21], enhanced vegetation index (EVI) [22], transformed vegetation index (TVI) [23], and the green normalized difference vegetation index (GNDVI) [24] derived from the AVIRIS-NG and Sentinel-2 images. Combining VIs and SIs to map bare soil minimizes vegetation influence and maximizes bare-soil features in order to extract the pixels most likely to be bare. Also, adding VIs could leverage the synergistic relationships between vegetation and soil properties, thereby improving the accuracy and robustness of the bare-soil mapping models. Agone and Bhamare [25] linked NDVI values of 0.00 to 0.20 with bare areas and several authors have used these VIs to classify bare-soil classes [26–29].

To verify the performance of our proposed index, we added a set of SIs to differentiate between bare soil, vegetation, and urban areas, including the bare soil index (BSI) [11]), brightness index (BI) [30], and normalized difference soil index 2 (NDSI2) [31]. These SIs were constructed with wavelength combinations from VIS to SWIR regions with the goal of differentiating between bare-soil status and vegetation status.

Table 1. Existing and new spectral indices used as covariates to map bare soil. For broadband indices, spectral bands were averaged to represent NIR (750–850 nm), red (600–700 nm), green (500–600 nm), blue (400–500 nm), and SWIR2.

Index	Equation
Vegetation Indices (VI)	
NDVI	$\frac{Red - NIR}{Red + NIR}$
MSAVI2	$\frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2}$
SAVI	$\frac{(NIR - Red)(1 + L)}{NIR - Red + L}$ where $L = 1$
EVI	$Green * \frac{(NIR - Red)}{NIR + C_1 * Red - C_2 * Blue + L}$ where $C_1 = 6, C_2 = 7.5, L = 1$
TVI	$\sqrt{\frac{NIR - Red}{NIR + Red} + 0.5} * 100$
GNDVI	$\frac{NIR - Green}{NIR + Green}$
Soil Indices (SI)	
BSI	$\sqrt{\frac{SWIR2 - Green}{SWIR2 + Green}} * 100$
BI	$\sqrt{\frac{Red^2 + Green^2}{2}}$
NDSI2	$\frac{SWIR2 - Green}{SWIR2 + Green}$
HBSI	$\frac{(SWIR2 + Green) - (NIR + Blue)}{(SWIR2 + Green) + (NIR + Blue)}$

2.4. Hyperspectral Bare Soil Index (HBSI)

We formulated the HBSI in Equation (1) by utilizing the spectral bands in the VIS, NIR, and SWIR2 regions. The HBSI takes advantage of the absorption bands at 400 to 500 nm (blue and green), the minimal absorption region around 900 nm (NIR), and the weak absorption near 2200 and 2300 nm (SWIR2). The VIS-NIR absorption in soil is mainly due to electronic transitions of the main active components of Fe-oxide minerals that do not have full d-orbitals [32]. Goethite (α -FeOOH) and hematite (α -Fe₂O₃) are the most common Fe-oxide minerals that exhibit broad absorption bands in the VIS-NIR regions. In the SWIR2, clay minerals and soil organic matter often show in narrow absorption spectral features [33,34]. Soil minerals that exhibit strong spectral maxima and minima of the second derivative in the SWIR2 region are gibbsite (Gbs) at around 2265 nm to 2285 nm, illite (Ill) at around 2205 nm to 2280 nm, and calcite (Cal) at around 2342 nm to 2367 nm [35].

$$HBSI = \frac{(SWIR2 + Green) - (NIR + Blue)}{(SWIR2 + Green) + (NIR + Blue)} \quad (1)$$

The spectral bands were averaged to represent the regions of green (500–600 nm), blue (400–500 nm), NIR (750–850 nm), and SWIR2 (~2080–2350 nm). With the HBSI, the reflectance spectra are normalized to allow a quantitative comparison between absorption features, such as depth, depth position, area of the absorption, etc., and to nondestructively retrieve bare soil [36].

2.5. Machine-Learning Classification Algorithms

We used random forest (RF) and support vector machine (SVM) classification algorithms to map bare soil [37] based on their proven performance and ability to handle complex classification tasks. We compiled a number of codes in R to run both algorithms [38,39]. RF and SVM are popular classifiers for digital soil-mapping using remote-sensing data to effectively capture complex relationships between spectral information and landcover classes, making it suitable for bare-soil detection [40,41]. RF and SVM provide a way to

select important covariates based on changes in the prediction accuracy when variables are added or deleted from models.

RF is a non-parametric supervised classifier that uses classification and regression tree (CART) through bagging, where it randomly picks a set of features and creates a classifier with a bootstrapped sample of the training data to grow a tree [35]. With RF training data selection, it is possible for the same sample to be picked several times, whereas others may not be picked at all. Apart from RF being quite robust with highly collinear variables, the random selection process at each tree node causes a low correlation among the trees and avoids over-fitting [42]. SVM is also a non-parametric supervised classifier used for pattern recognition, classification, and regression analysis. SVM is robust when processing a small number of training samples, but efficient at producing accurate maps when applied as a classifier to satellite images, as reported by Mountrakis et al. [43].

2.6. Covariates, Training, and Test Datasets

To construct the RF and SVM algorithms for mapping bare soil, we used the image indices from AVIRIS-NG and Sentinel-2 as covariates. For each algorithm, we ran the classification algorithms using two sets of covariates: one set that contained all six VIs and another set with all four SIs (Table 1). We also ran a model ensemble that combined the RF and SVM algorithms. Afterwards, we ran the classification models using only the top three and top two important VI and SI variables from the previous step, respectively. We ran the models for the last time using only the two most important VI and SI covariates from the previous step.

To produce the bare-soil maps, we limited the class features to only three: bare soil, urban areas, and others (forest, vegetation, etc.). We extracted high-quality samples from the study area using Google Earth, and each class label was visually interpreted using spectral profiles and using our expert knowledge of the field sites. These samples served as regions of interest (ROI) for the RF and SVM classification process. Each ROI was assigned to a specific LULC class. A total of 540 polygon samples were extracted: 200 urban, 200 bare soil, and 140 others. We then split the samples into training and testing sets. We used a splitting criterion of 70–30, where 70% of the sample data were used for calibration and 30% for model validation.

To evaluate the effectiveness of the classification models, we generated a classification error matrix. We utilized conventional accuracy metrics, including overall accuracy (OA) and kappa statistics (KA) to quantify the performance of the RF and SVM models. These metrics provided a comprehensive assessment of the models' ability to accurately classify bare soil, urban areas, and other landcover categories. By employing these indicators, we aimed to thoroughly evaluate the performance of the RF and SVM models in accurately mapping bare soil in the study area. The utilization of established accuracy metrics allowed for a reliable assessment of the performance of the models, enabling us to validate their efficacy in producing accurate bare-soil maps.

3. Results

Figure 2 shows the final classified bare-soil maps derived using an ensemble of RF and SVM from (a) AVIRIS-NG (4-m spatial resolution) and (b) Sentinel-2 (10-m spatial resolution) images. The map using Sentinel-2 shows that more areas were classified as bare soil (149 km²) compared to the classification map using AVIRIS-NG (124 km²) images. The magnified section (Figure 2c) shows the delineation of the bare-soil class from urban areas. Furthermore, the black box highlights the edges of a patch of urban areas misclassified as bare soil.

Table 2 lists the top important variables according to the RF and SVM classification models for each set of variables and images. For Sentinel-2 and AVIRIS-NG, NDVI and TVI were predominant variables for RF and SVM, ranking mostly in the top two in six for the VI set. The HBSI and NDSI2 were the most important variables for the SI set. When the important VI and SI variables were combined, HBSI and NDVI ranked first. Between the

NDVI and HBSI, the latter was the variable with a higher classification importance in both the RF and SVM models and images.

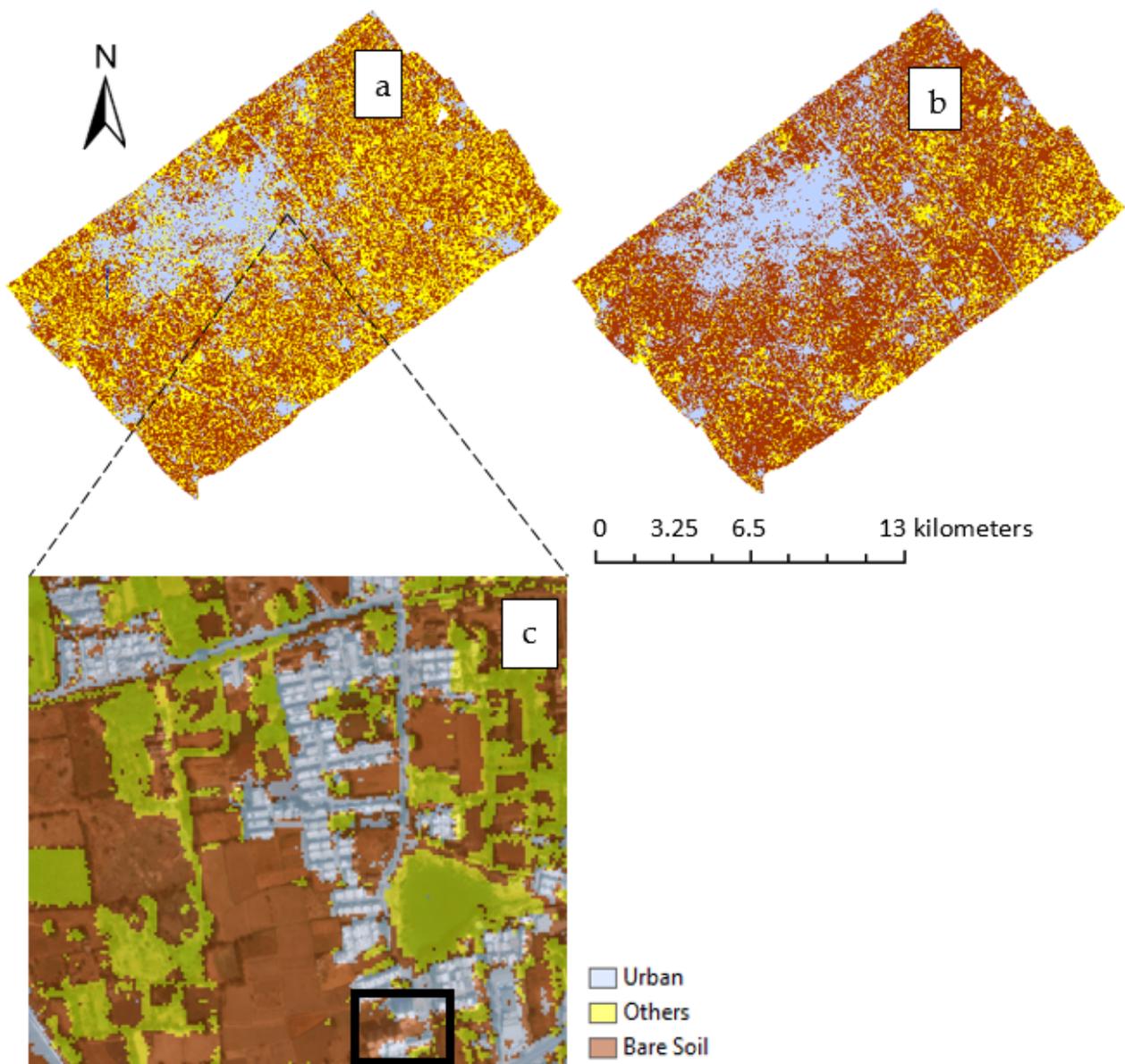


Figure 2. Bare soil-urban map derived using an ensemble of RF and SVM from: (a) AVIRIS-NG; (b) Sentinel-2 images; and (c) magnified section to highlight the separation of classes.

Table 2. Rankings of overall variable importance for RF and SVM models using different sets of covariates. The symbol * means that the variable does not have any contribution.

Variable Set	Sentinel-2		AVIRIS-NG	
	RF	SVM	RF	SVM
6 VIs	1. NDVI	1. TVI	1. NDVI	1. TVI
	2. TVI	2. NDVI	2. GNDVI	2. NDVI
	3. GNDVI	3. SAVI	3. TVI	3. GNDVI
	4. EVI	4. GNDVI	4. SAVI	4. SAVI
	5. SAVI	5. MSAVI	5. EVI	5. MSAVI
	6. MSAVI	6. EVI	6. * MSAVI	6. * EVI

Table 2. Cont.

Variable Set	Sentinel-2		AVIRIS-NG	
	RF	SVM	RF	SVM
4 SIs	1. HBSI	1. HBSI	1. HBSI	1. HBSI
	2. NDSI2	2. NDSI2	2. NDSI2	2. NDSI2
	3. BSI	3. BSI	3. BI	3. BSI
	4. BI	4. BI	4. BSI	4. BI
3 VIs & 2 SIs	1. HBSI	1. NDVI	1. NDVI	1. HBSI
	2. NDVI	2. HBSI	2. HBSI	2. NDSI2
	3. NDSI2	3. NDSI2	3. GNDVI	3. NDVI
	4. TVI	4. GNDVI	4. NDSI2	4. TVI
	5. GNDVI	5. TVI	5. TVI	5. GNDVI
NDVI & HBSI	1. HBSI	1. HBSI	1. HBSI	1. HBSI
	2. NDVI	2. NDVI	2. NDVI	2. NDVI

The Sentinel-2 overall validation accuracy from the final dataset with NDVI and HBSI were 93.5% (K = 87.8%), 94.9% (K = 90%), and 93.6% (K = 87.9%) for RF, SVM, and the ensemble, respectively (Table 3). For AVIRIS-NG, the OA and K values were higher than the Sentinel-2 at 98.6% (K = 97.4%), 98.5% (K = 97.3%), and 98.6% (K = 97.4%) for RF, SVM, and the ensemble, respectively. Also, the final combination of NDVI and HBSI saw a slight improvement in overall accuracy for training and validation when compared to the other sets of covariates (e.g., 6 VIs, 4 SIs, 3 VIs and 2 SIs) for both Sentinel-2 and AVIRIS-NG images.

Table 3. Tabulated accuracy statistics for RF, SVM, and their ensemble for the different sets of covariates. All *p*-values are less than 0.001. OA = overall accuracy (%), K = kappa (%).

Variable Set	Sentinel-2				AVIRIS-NG			
	Training		Validation		Training		Validation	
	OA	K	OA	K	OA	K	OA	K
6 VIs								
RF	92.4	89.6	92.3	89.2	97.9	96.7	97.9	96.7
SVM	91.6	87.3	91.8	87.4	97.1	95.7	97.0	95.5
Model Ensemble	92.0	88.9	91.8	87.5	97.3	96.1	97.0	96.9
4 SIs								
RF	90.6	89.7	90.5	89.9	98.2	97.2	98.5	97.1
SVM	91.4	88.6	91.9	87.7	97.1	96.2	97.3	96.2
Model Ensemble	90.3	87.9	90.5	87.2	97.7	97.0	98.1	97.1
3 VIs & 2 SIs								
RF	91.2	87.9	90.3	88.1	98.2	97.3	98.7	97.2
SVM	91.5	88.2	92.4	87.7	95.7	94.6	95.7	93.5
Model Ensemble	91.4	88.0	91.9	87.1	96.8	95.9	96.7	95.2
NDVI & HBSI								
RF	93.9	88.1	93.5	87.8	98.2	97.3	98.6	97.4
SVM	94.0	89.3	94.9	90.0	98.2	97.4	98.5	97.3
Model Ensemble	92.0	87.7	93.6	87.9	98.2	97.4	98.6	97.4

A matrix of class accuracies using the final model of NDVI and HBSI (Table 4) are shown in terms of the Kappa coefficient and overall accuracy. For Sentinel-2, the overall accuracies were 93.5%, 94.9%, and 93.6% for RF, SVM, and ensemble models, respectively. The kappa values range from 87.8 to 90.0, which indicate an acceptable level of accuracy of the classified maps. However, for the AVIRIS-NG, the overall accuracies were higher at 98.6%, 98.5%, and 98.6%, for RF, SVM, and ensemble models, respectively. The range of kappa values was also higher compared to Sentinel-2, with a range of 97.3 to 97.4.

Table 4. Confusion matrix from validation dataset using the final model of NDVI and HBSI.

	Sentinel-2								
	RF			SVM			Ensemble		
	Soil	Urban	Others	Soil	Urban	Others	Soil	Urban	Others
Soil	1324	92	13	1393	78	33	1317	82	13
Urban	96	2947	2	88	2971	2	106	2957	2
Others	93	23	376	32	13	256	90	23	376
Overall	93.5			94.9			93.6		
Kappa	87.8			90.0			87.9		
AVIRIS-NG									
Soil	8744	45	173	8636	72	20	8744	47	175
Urban	150	2155	18	104	2080	5	150	2153	18
Others	3	1	17,237	157	49	17,403	3	1	17,235
Overall	98.6			98.5			98.6		
Kappa	97.4			97.3			97.4		

Table 5 lists the accuracy statistics after we ran the RF-SVM ensemble model using the individual soil indices. Similar to the results in Table 2, the HBSI and NDSI2 were the two indices with over 90% in overall accuracy, for both Sentinel-2 and AVIRIS-NG images. The BI had the lowest OA overall.

Table 5. Tabulated accuracy statistics for the individual soil index using an ensemble of RF and SVM models. All *p*-values are less than 0.001. OA = overall accuracy (%), K = kappa (%).

Soil Index	Sentinel-2				AVIRIS-NG			
	Training		Validation		Training		Validation	
	OA	K	OA	K	OA	K	OA	K
BSI	89.4	88.5	89.2	88.8	92.3	90.7	92.1	90.9
BI	88.6	86.1	88.2	86.3	89.3	88.6	89.5	88.7
NDSI2	90.0	88.4	90.1	87.7	93.7	92.3	92.9	91.8
HBSI	91.3	88.6	91.7	87.7	94.2	94.4	94.1	93.6

4. Discussion

4.1. Characteristics of HBSI vs. Other Indices

The results of our study indicate that the HBSI outperformed other existing indices for mapping bare soil. In our analysis, we observed that the HBSI, which is calculated using the blue, green, NIR, and SWIR2 spectral bands, exhibited unique advantages compared to other indices.

One distinctive characteristic of the HBSI is its utilization of the blue band in addition to the NIR and SWIR2 bands. These bands demonstrated the capability of the HBSI to discriminate bare-soil features from urban and other land-use classes. Unlike other bare-soil indices that only utilize SWIR2 and green bands [11,30], the HBSI took advantage of the reflectance of the blue band that allows for the capture of additional information. Furthermore, by combining the blue band with the NIR and then normalizing the difference, the dissimilarities between the strongest and weakest features within these spectral regions were emphasized. A similar study by Liu et al. [44] also highlighted the SWIR and the blue bands as being the two important spectral bands for bare-soil mapping since they represent the highest and the lowest reflectance for soil, respectively.

The HBSI is a promising alternative to existing bare-soil indices as it dramatically widens the gap between bare soil and other classes based on their unique spectral features. When used alone, the HBSI showed its effectiveness in bare-soil mapping with over 91% accuracy for Sentinel-2 and AVIRIS-NG images.

Finding the best combination of indices is an important step towards achieving the efficiency needed for mapping and understanding soil behavior. The combination of HBSI and NDVI showed a slightly better performance in bare-soil classification, with >92% for Sentinel-2 images and >97% for AVIRIS-NG images. The combination of indices minimized vegetation influence and maximized bare-soil features, which resulted in positive classification for bare-soil pixels only [15]. These VIs were used by several authors to classify bare soil, either separately or in combination [26–29]. A study in Italy using Sentinel-2 combined NDVI and BSI and delivered good discrimination between bare soil and other land classes [12].

4.2. Limitations of HBSI

While the HBSI has been proven useful for Sentinel-2 and AVIRIS-NG images in this study and it shows promise, it is important to acknowledge some limitations and uncertainties associated with the results. Since the HBSI uses the SWIR2 band (~2080–2350 nm), the potential of the index should be validated with other satellite images that contain SWIR2 and with broader spatial resolutions, such as Landsat 8/9. Second, since the spectral behavior of bare soil can vary with different soil types or landscapes [45], the HBSI could behave differently when applied to mudflats or dunes. We only tested the HBSI in a study area that is considered as farmland, with diverse agricultural management practices. More tests on other soil landscapes are needed, and caution should be exercised when applying the HBSI to different regions. Third, this study did not include a temporal analysis of bare soil. Seasonal soil mapping is needed to differentiate between bare lands that are fallow agriculture fields from those that are construction sites [15]. Fourth, because the HBSI relies on the SWIR2 wavelength, the index may not be useful for unmanned aerial vehicle (UAV) or drone images, for which SWIR is often unavailable. Lastly, although the combination of the HBSI and NDVI improved classification accuracy, there may still be cases where misclassifications occur, particularly in complex landscapes, or patches of urban areas (Figure 2c), or under challenging conditions. Future research could focus on addressing these limitations to further improve the accuracy and robustness of bare-soil mapping using the HBSI.

4.3. Performance of Ensemble Model

The effectiveness of an ensemble model depends on the precision of the individual models and conditional bias in simulated values during model training [46]. Machine learning techniques enable automated feature selection and extraction, identifying the most relevant spectral bands or indices for mapping bare soil. By focusing on informative features, these machine learning models could improve the accuracy and efficiency of the mapping process [40]. Combining machine learning models into a single ensemble produced a more representative soil map since it highlighted the agreement across algorithms [47]. In our study, the RF-SVM ensemble surpassed the performance of the individual models. In other words, the ensemble was even stronger as a result of the rather high performance of the individual models, despite the fact that prior research described the ensemble algorithm as having limited interpretability [48]. We may test additional machine learning models in the future, add ensemble criteria to exclude underperformers, and broaden the sorts of models chosen to boost ensemble performance.

4.4. HBSI Potential Area of Focus for Future Research

There are several ways in which future research can improve the HBSI and its application in bare-soil mapping, such as the fusion of multi-temporal data. Bare soil conditions can vary over time due to changes in season and land management practices, and due to natural disturbances [15]. By integrating multi-temporal hyperspectral data, we will be able to capture temporal patterns and enhance the performance of the HBSI for mapping bare soil. Time series analysis, data-fusion techniques, and change-detection algorithms could be applied to utilize multi-temporal data and improve mapping accuracy. Second, the

integration of ancillary data [29], such as soil moisture measurements, terrain information, or soil-texture data, could also improve the accuracy and reliability of bare-soil mapping using the HBSI. Machine learning techniques could be utilized to effectively combine and exploit these additional data sources.

5. Conclusions

In this paper, we developed a bare-soil spectral index called the HBSI and investigated its effectiveness using two machine learning algorithms applied to two different satellite images. The statistical tests showed that the HBSI performs better when compared with other bare-soil indices and could meet the requirements for bare-soil classification. Mapping bare soil from Sentinel-2 (multispectral) and AVIRIS-NG (hyperspectral) images, as proposed in this study, is feasible and reliable. We believe in the potential of HBSI to improve the accuracy of bare-soil remote-sensing mapping.

Author Contributions: Conceptualization, E.A.L.S.; methodology, E.A.L.S.; software, S.S.K.; validation, E.A.L.S.; formal analysis, E.A.L.S.; investigation, E.A.L.S.; resources, S.S.K.; data curation, E.A.L.S.; writing—original draft preparation, E.A.L.S.; writing—review and editing, S.S.K.; visualization, E.A.L.S.; supervision, S.S.K.; project administration, S.S.K.; funding acquisition, S.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Aeronautics and Space Administration (Grant number 80NSSC17K0653 P00001) for the joint NASA and Indian Space Research Organization AVIRIS-NG Campaign in India. The study was also supported by NIFA/USDA through Central State University Evans-Allen Research Program Fund Number NI201445XXXXG018-0001.

Data Availability Statement: Not applicable.

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

References

1. Biancari, L.; Aguiar, M.R.; Cipriotti, P.A. Grazing Impact on Structure and Dynamics of Bare Soil Areas in a Patagonian Grass-Shrub Steppe. *J. Arid Environ.* **2020**, *179*, 104197. [\[CrossRef\]](#)
2. Wythers, K.; Lauenroth, W.; Paruelo, J. Bare-Soil Evaporation Under Semiarid Field Conditions. *Soil Sci. Soc. Am. J.-SSSAJ* **1999**, *63*, 1341–1349. [\[CrossRef\]](#)
3. Lehmann, P.; Merlin, O.; Gentine, P.; Or, D. Soil Texture Effects on Surface Resistance to Bare-Soil Evaporation. *Geophys. Res. Lett.* **2018**, *45*, 10–398. [\[CrossRef\]](#)
4. Li, J.; Okin, G.S.; Skiles, S.M.; Painter, T.H. Relating Variation of Dust on Snow to Bare Soil Dynamics in the Western United States. *Environ. Res. Lett.* **2013**, *8*, 044054. [\[CrossRef\]](#)
5. Almazroui, M.; Mashat, A.; Assiri, M.E.; Butt, M.J. Application of Landsat Data for Urban Growth Monitoring in Jeddah. *Earth Syst. Env.* **2017**, *1*, 25. [\[CrossRef\]](#)
6. He, T.; Gao, F.; Liang, S.; Peng, Y. Mapping Climatological Bare Soil Albedos over the Contiguous United States Using MODIS Data. *Remote Sens.* **2019**, *11*, 666. [\[CrossRef\]](#)
7. Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The Use of Remote Sensing in Soil and Terrain Mapping—A Review. *Geoderma* **2011**, *162*, 1–19. [\[CrossRef\]](#)
8. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Chapter Four—Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Cambridge, MA, USA, 2015; Volume 132, pp. 139–159.
9. Comstock, J.P.; Sherpa, S.R.; Ferguson, R.; Bailey, S.; Beem-Miller, J.P.; Lin, F.; Lehmann, J.; Wolfe, D.W. Carbonate Determination in Soils by Mid-IR Spectroscopy with Regional and Continental Scale Models. *PLoS ONE* **2019**, *14*, e0210235. [\[CrossRef\]](#)
10. Goudge, T.A.; Russell, J.M.; Mustard, J.F.; Head, J.W.; Bijaksana, S. A 40,000 Yr Record of Clay Mineralogy at Lake Towuti, Indonesia: Paleoclimate Reconstruction from Reflectance Spectroscopy and Perspectives on Paleolakes on Mars. *GSA Bull.* **2017**, *129*, 806–819. [\[CrossRef\]](#)
11. Wentzel, K. Determination of the Overall Soil Erosion Potential in the Nsikazi District (Mpumalanga Province, South Africa) Using Remote Sensing and GIS. *Can. J. Remote Sens.* **2002**, *28*, 322–327. [\[CrossRef\]](#)
12. Mzid, N.; Pignatti, S.; Huang, W.; Casa, R. An Analysis of Bare Soil Occurrence in Arable Croplands for Remote Sensing Topsoil Applications. *Remote Sens.* **2021**, *13*, 474. [\[CrossRef\]](#)
13. Lin, H.; Wang, J.; Liu, S.; Qu, Y.; Wan, H. Studies on Urban Areas Extraction from Landsat TM Images. In Proceedings of the Proceedings 2005 IEEE International Geoscience and Remote Sensing Symposium, Seoul, Republic of Korea, 29–29 July 2005; IGARSS '05; Volume 6, pp. 3826–3829.

14. Koroleva, P.V.; Rukhovich, D.I.; Rukhovich, A.D.; Rukhovich, D.D.; Kulyanitsa, A.L.; Trubnikov, A.V.; Kalinina, N.V.; Simakova, M.S. Location of Bare Soil Surface and Soil Line on the RED–NIR Spectral Plane. *Eurasian Soil Sc.* **2017**, *50*, 1375–1385. [[CrossRef](#)]
15. Li, H.; Wang, C.; Zhong, C.; Su, A.; Xiong, C.; Wang, J.; Liu, J. Mapping Urban Bare Land Automatically from Landsat Imagery with a Simple Index. *Remote Sens.* **2017**, *9*, 249. [[CrossRef](#)]
16. He, C.; Shi, P.; Xie, D.; Zhao, Y. Improving the Normalized Difference Built-up Index to Map Urban Built-up Areas Using a Semiautomatic Segmentation Approach. *Remote Sens. Lett.* **2010**, *1*, 213–221. [[CrossRef](#)]
17. Zhao, D.; Arshad, M.; Li, N.; Triantafylis, J. Predicting Soil Physical and Chemical Properties Using VIS–NIR in Australian Cotton Areas. *Catena* **2021**, *196*, 104938. [[CrossRef](#)]
18. Zhao, D.; Arshad, M.; Wang, J.; Triantafylis, J. Soil Exchangeable Cations Estimation Using Vis–NIR Spectroscopy in Different Depths: Effects of Multiple Calibration Models and Spiking. *Comput. Electron. Agric.* **2021**, *182*, 105990. [[CrossRef](#)]
19. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
20. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
21. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
22. Liu, H.Q.; Huete, A. A Feedback Based Modification of the NDVI to Minimize Canopy Background and Atmospheric Noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [[CrossRef](#)]
23. McDaniel, K.C.; Haas, R.H. Assessing Mesquite-Grass Vegetation Condition from Landsat. *Photogramm. Eng. Remote Sens.* **1982**, *48*, 441450.
24. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
25. Agone, V.; Bhamare, S. Change Detection of Vegetation Cover Using Remote Sensing and GIS. *J. Res. Dev.* **2012**, *2*, 91–102.
26. Singh, R.G.; Engelbrecht, J.; Kemp, J. Change Detection of Bare Areas in the Xolobeni Region, South Africa Using Landsat NDVI. *South Afr. J. Geomat.* **2015**, *4*, 138–148. [[CrossRef](#)]
27. Phinzi, K.; Ngetar, N.S. Mapping Soil Erosion in a Quaternary Catchment in Eastern Cape Using Geographic Information System and Remote Sensing. *South Afr. J. Geomat.* **2017**, *6*, 11–29. [[CrossRef](#)]
28. Sepuru, T.K.; Dube, T. An Appraisal on the Progress of Remote Sensing Applications in Soil Erosion Mapping and Monitoring. *Remote Sens. Appl. Soc. Environ.* **2018**, *9*, 1–9. [[CrossRef](#)]
29. Hamzehpour, N.; Shafizadeh-Moghadam, H.; Valavi, R. Exploring the Driving Forces and Digital Mapping of Soil Organic Carbon Using Remote Sensing and Soil Texture. *Catena* **2019**, *182*, 104141. [[CrossRef](#)]
30. Metternicht, G.; Zinck, J.A. Spatial Discrimination of Salt- and Sodium-Affected Soil Surfaces. *Int. J. Remote Sens.* **1997**, *18*, 2571–2586. [[CrossRef](#)]
31. Deng, Y.; Wu, C.; Li, M.; Chen, R. RNDISI: A Ratio Normalized Difference Soil Index for Remote Sensing of Urban/Suburban Environments. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 40–48. [[CrossRef](#)]
32. Tappert, M.; Rivard, B.; Giles, D.; Tappert, R.; Mauger, A. Automated Drill Core Logging Using Visible and Near-Infrared Reflectance Spectroscopy: A Case Study from the Olympic Dam IOCG Deposit, South Australia. *Econ. Geol.* **2011**, *106*, 289. [[CrossRef](#)]
33. Curcio, D.; Ciraolo, G.; D’Asaro, F.; Minacapilli, M. Prediction of Soil Texture Distributions Using VNIR–SWIR Reflectance Spectroscopy. *Procedia Environ. Sci.* **2013**, *19*, 494–503. [[CrossRef](#)]
34. Zhao, D.; Triantafylis, J.; Zhao, X. A Vis–NIR Spectral Library to Predict Clay in Australian Cotton Growing Soil. *Soil Sci. Soc. Am. J.* **2018**, *82*, 1347–1357. [[CrossRef](#)]
35. Mendes, W.D.S.; Demattê, J.A.M.; Bonfatti, B.R.; Resende, M.E.B.; Campos, L.R.; Costa, A.C.S. da A Novel Framework to Estimate Soil Mineralogy Using Soil Spectroscopy. *Appl. Geochem.* **2021**, *127*, 104909. [[CrossRef](#)]
36. Chabrilat, S.; Goetz, A.F.H.; Krosley, L.; Olsen, H.W. Use of Hyperspectral Images in the Identification and Mapping of Expansive Clay Soils and the Role of Spatial Resolution. *Remote Sens. Environ.* **2002**, *82*, 431–445. [[CrossRef](#)]
37. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
38. FAO. *Soil Organic Carbon Mapping Cookbook*, 2nd ed.; FAO: Rome, Italy, 2018; ISBN 978-92-5-130440-2.
39. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
40. Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; Schmidt, M.G. An Overview and Comparison of Machine-Learning Techniques for Classification Purposes in Digital Soil Mapping. *Geoderma* **2016**, *265*, 62–77. [[CrossRef](#)]
41. Wadoux, A.; Samuel-Rosa, A.; Poggio, L.; Mulder, V.L. A Note on Knowledge Discovery and Machine Learning in Digital Soil Mapping. *Eur. J. Soil Sci.* **2019**, *71*, 133–136. [[CrossRef](#)]
42. Salas, E.A.L.; Subburayalu, S.K. Modified Shape Index for Object-Based Random Forest Image Classification of Agricultural Systems Using Airborne Hyperspectral Datasets. *PLoS ONE* **2019**, *14*, e0213356. [[CrossRef](#)]
43. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [[CrossRef](#)]
44. Liu, Y.; Meng, Q.; Zhang, L.; Wu, C. NDBSI: A Normalized Difference Bare Soil Index for Remote Sensing to Improve Bare Soil Mapping Accuracy in Urban and Rural Areas. *Catena* **2022**, *214*, 106265. [[CrossRef](#)]

45. Ge, Y.; Thomasson, J.A.; Sui, R. Remote Sensing of Soil Properties in Precision Agriculture: A Review. *Front. Earth Sci.* **2011**, *5*, 229–238. [[CrossRef](#)]
46. Sylvain, J.-D.; Anctil, F.; Thiffault, É. Using Bias Correction and Ensemble Modelling for Predictive Mapping and Related Uncertainty: A Case Study in Digital Soil Mapping. *Geoderma* **2021**, *403*, 115153. [[CrossRef](#)]
47. Taghizadeh-Mehrjardi, R.; Minasny, B.; Toomanian, N.; Zeraatpisheh, M.; Amirian-Chakan, A.; Triantafyllis, J. Digital Mapping of Soil Classes Using Ensemble of Models in Isfahan Region, Iran. *Soil Syst.* **2019**, *3*, 37. [[CrossRef](#)]
48. Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. Soil Organic Carbon Concentrations and Stocks on Barro Colorado Island—Digital Soil Mapping Using Random Forests Analysis. *Geoderma* **2008**, *146*, 102–113. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.