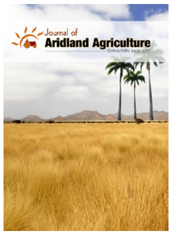
Journal of Aridland Agriculture 2023, 9: 81-91
doi: 10.25081/jaa.2023.v9.8566
https://updatepublishing.com/journal/index.php/jaa

Research Article



ISSN: 2455-9377

Classification of Crops through Self-Supervised Decomposition for Transfer Learning

J. Jayanth[1]*, H. K. Ravikiran[2], K. M. Madhu[3]

[1]Department of Electronics and Communication Engineering, GSSS Institute of Engineering & Technology for Women, Mysore-570016, Karnataka, India, [2]Department of Electronics and Communication Engineering, Navkis College of Engineering, Hassan-573217, Karnataka, India, [3]Department of Civil Engineering, Rajeev Institute of Technology, Hassan-573201, Karnataka, India

ABSTRACT

The 2S-DT (Self-Supervised Decomposition for Transfer Learning) model, created for crop categorization using remotely sensed data, is a unique method introduced in this paper. It deals with the difficulty of incorrectly identifying crops with comparable phenology patterns, a problem that frequently arises in agricultural remote sensing. Two datasets from Nanajangudu taluk in the Mysore district, which has a widely varied irrigated agriculture system, are used to assess the model. Using self-supervised learning, the 2S-DT model addresses the misclassification issue that frequently occurs when working with unlabeled classes, especially in high-resolution images. It uses class decomposition (CD) layer and a downstream learning approach. Using the model's learning and the particulars of each geographical context, this layer improves the information's arrangement. Our model architecture's foundation is ResNet, a well-known deep learning framework. Each residual block in our ResNet architecture is made up of two 3x3 convolutional layers. Each convolutional layer is followed by batch normalization and Rectified Linear Unit (ReLU) activation functions, which improve the model's capacity for learning. We utilized a 7x7 convolutional layer with 64 filters and a stride of 2 for Conv1 in ResNet18, resulting in an output size of 112x112x64. Conv2, which consists of Res2a and Res2b, generated an output with the dimensions 48x48x64. Conv3, which included Res3a and Res3b, produced an output with the dimensions 28x28x128. These architectural selections were made with our experimental needs in mind. The 2S-DT model's newly added features make it easier to identify classes and update weights, improving the stability of the features' spatial and spectral data. Extensive tests performed on two datasets show the model's viability. Overall accuracy has improved significantly, with the 2S-DT model surpassing comparable models like TVSM, 3DCAE, and GAN Model by obtaining 95.65% accuracy for dataset 1 and 88.91% accuracy for dataset 2.

KEYWORDS: Self-Supervised (2S) Transfer Learning, High-Resolution Image Classification, Spectral Features, TVSM, 3DCAE, GAN Model

Received: August 21, 2023
Revised: September 23, 2023
Accepted: October 3, 2023
Published: October 17, 2023

***Corresponding Author:**
J. Jayanth
E-mail: ravikiranhsn@gmail.com

INTRODUCTION

The potential use of remote sensing in the classification of crops over a broad area has been broadly investigated on the basis of the classification and mapping of croplands (Arel *et al.*, 2010; Radford *et al.*, 2015). Using remote sensing data, the Department of Agriculture and Cooperation (DAC), Ministry of Agriculture, Govt. of India, initiated steps to set up a center for routine check-up of crop statistics using AWIFS and LISS III data 20. The Mahalanobis National Crop Forecast Centre (MNCFC) was set up by the Govt. of India, New Delhi, for estimating the crop yield and its planting area using land use (LU)/Land Cover (LC) data.

Initially, high-resolution RS data such as LISS IV, PAN, Landsat 8, and Sentinel-2 act as the main data source for information on crop area (Bolton & Friedl, 2013; Esch *et al.*, 2014; Gao *et al.*, 2017). As is for the most part the case with measurable testing, the more preparation sets that are not entirely settled, the more noteworthy the probability of getting the right characterization exactness; this assumption is also true with MLC. Parametric classifiers fail to classify when there is insufficient training data and when they are unable to satisfy the rule of thumb defined for training data set size (Gallego *et al.*, 2012; Hedhli *et al.*, 2016).

Deep learning-based pixel-wise classifiers have acquired consideration in RS data classification (Kussul *et al.*, 2017). Even though a nonparametric classifier algorithm's accuracy is less compared to TVSM, RNN, 2D TVSM and others, the main disadvantage associated with them is that they are either expensive in computation or complex in execution because of the prerequisite of different parameter settings for their ideal exhibition, which is widely used in remote sensing-based crop classification (Mathur & Foody, 2008; Löw *et al.*, 2013; LeCun *et al.*, 2015; Li *et al.*, 2017). The black-box nature of the ANN structure and the knowledge automation problem associated with fuzzy systems in soft classification have failed to prove themselves user-friendly and compelled the analysts to inevitably use MLC on non-Gaussian data (Mathur & Foody, 2008; Omkar *et al.*, 2008; Mei *et al.*, 2019). A review of machine learning classification methods was proposed by Kussul N (Bruzzone *et al.*, 2005; Kussul *et al.*, 2017; Jayanth *et al.*, 2020), one of their conclusions is that there is typically no valid or right methodology, yet it is important to consider approaches for choosing a suitable strategy for a specified issue.

In addition, this transformation-based model has been used to resolve misclassification in similar crop phenology by selecting the features of targeted classes and improving the accuracy of the classification. When evolutionary strategic algorithms are advocated, the Spare learning and deep belief networks do gain importance due to their transformational nature used for classification. Even though GAN (Radford *et al.*, 2015) an unsupervised learning method, achieves better results in overall classification accuracy (OCA) through hyperspectral data, it fails to update the velocity of each particle during its parameter settings to achieve optimal performance. A deep residual network with 49 layers was used in Deep Multi-View Learning algorithm to extract features for classification and also to overcome the labeled training sample in a sequential manner. 3DCAE takes much longer to classify each class than other methods (Mei *et al.*, 2019).

In this work, we use self-supervised decomposition for the transfer learning model, which can overcome the difficulties of keeping track of subpixel heterogeneity without aligning information while testing and training the RS data and for validation of classified data. The objective of this article is to classify multispectral LISS IV remote sensing data to classify agricultural land cover and assess the 2S-DT technique by comparing it with other approaches. For this purpose, we have used 2S-DT to urge coarse-to-fine exchange learning in light of a self-managed test deterioration approach. 2S-DT can manage any abnormalities in the information dispersion and the restricted accessibility of preparing tests in certain classes. The commitments of this work can be summed up as follows:

1. Give a clever instrument to self-directed example disintegration utilizing an enormous arrangement of unlabelled classes for an appearance preparation.
2. Give a nonexclusive coarse-to-fine exchange learning procedure to step by step work on the power of information change from enormous scope picture acknowledgment errands to a particular class arrangement.
3. Give a downstream CD layer in the downstream preparation stage that can adapt to any abnormalities in the information circulation and improve its neighbourhood.

The present article is organized in the following manner: An introduction to the article and the study area is elaborated upon in Section 1, while Section 2 provides a brief introduction to the self-supervised decomposition for Transfer learning (2S-DT) model. Sections 3 and 4 present the results and discussion, as well as the conclusion.

Dataset

Datasets drawn are from the DodakavalandeHobli, in Nanjangudu taluk, Mysore district, Karnataka, India. The data drawn are verified using cadastral maps and topo sheets using the official data provided by the Government of India and the Government of Karnataka. Dataset 1 consists of 42*776 pixels, and six classes are marked. The number of training and testing samples for Dataset 1 is shown in Table 1. The study area consists of three zones: residential, agricultural, and natural environment. Dataset 2's study area consists of four zones with 422*1056 pixels, and eleven classes are marked. The number of training and testing samples for Dataset 2 are tabulated in Table 2, in which coconut trees, some teak trees, and other trees are permanent and have covered 3 ha, 1ha is the residential, and the remaining 6 ha are used for agriculture, as shown in Figures 1 and 2. Plot level crop inventory was carried out using a Trimble GPS device in the field survey, where 720 reference plots were collected for the land cover information. There is a substantial variation in the dimensions of the crop plots and also in cultivation practices. In this work, the satellite data available for this work is from the month of November. Following that, we can notice rabi crops in November. We choose 6 observed surface classes as shown in Table 1 for Dataset1, 11 classes as shown in Table 2 for Dataset 2. The specifications of the image data product for this study area are shown in Table 3. The data are from the LISS-IV (Linear Imaging and Self Scanning) sensor, which was procured from the National Remote Sensing Centre (NRSC) in Hyderabad, India. Information from the satellite data is geo-referred and projected regarding the reference of global positioning system (GPS) readings (Table 4).

Table 1: Number of training and testing sample for Dataset1

S. No.	Class	Training	Testing
1	AnnualCrop	5	1024
2	FallowLand	5	421
3	WaterBodies	5	345
4	Built-upland	5	480
5	Natural Space	5	270
6	other	5	80
	Total	30	2,620

Table 2: Number of training and testing sample for Dataset2

S. No.	Class	Training	Testing
1	Jowar	1357	1024
2	Turmeric	4571	421
3	TurDal	4018	345
4	Bark	3258	480
5	Horsegram	2362	270
6	Vegetable	1362	80
7	Plantation	1978	24
8	Built-upland	1918	256
9	Fallowsland	1978	356
10	ShurbLand	1362	424
11	WaterBodies	1052	93
	Total	25,216	3,773

Table 3: Confusion matrix for TVSM algorithm for Dataset1

Classes	1	2	3	4	5	6	Row Total	UA%
1	518			10	1		529	97.92
2	6	16			1		23	69.56
3			17	30		6	53	32.07
4	19		2	118	2		141	83.68
5	8	2		10	39		59	66.10
6				10		60	70	85.17
Column Total	551	18	19	151	43	66	875	
PA %	94.01	88.88	89.47	78.14	90.69	90.90		OCA87.77
Kappa	0.783	0.73	0.78	0.65	0.89	0.70		

1-AnnualCrop, 2-Waterbodies, 3-Built-upland, 4-Fallowland, 5-NaturalSpace and 6-others

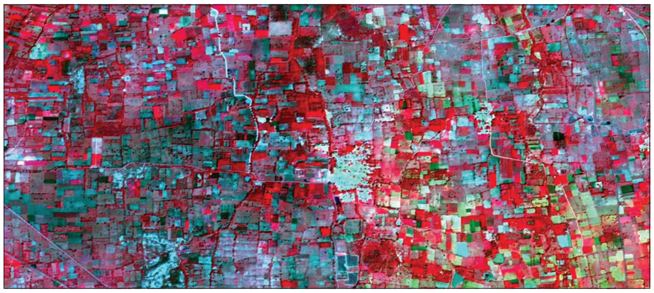


Figure 1: Study area of Dataset1 Devnur village

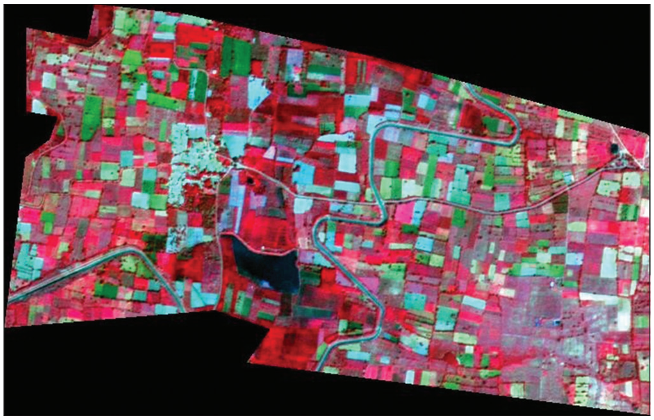


Figure 2: Study area of Dataset 2 Husguru village

PROPOSED SELF-SUPERVISED DECOMPOSITION FOR TRANSFER LEARNING (2S-DT) MODEL

In this section, the 2S-DT model for the classification of remote sensed data will be examined and analyzed. Starting with the architectural outline, the section discusses the workflow and formalizes the method illustrated in Figure 3.

The proposed model, 2S-DT, consists of three training phases.

Phase1

- Adopt the self-supervised algorithm to learn the pattern features of images like crops, Jowar, and building blocks.
- Preparing an AE model to remove the profound to conquer the restricted accessibility of marked images by utilizing the colossal accessibility of unlabeled
- When the component space of the unlabeled image dataset is built, a sample decomposition approach is used to make pseudo-marks for the classification.

Phase 2

We utilize the pseudo-marks to accomplish coarse exchange learning by utilizing an ImageNet-pretrained CNN model for the ordered (classification) pseudo-mark, which helps in fine-tuning the parameter.

Phase 3

Trained convolutional features have been utilized to accomplish the training of RS data. This task is more explicit by adjusting a fine exchange, gaining from unique features in the image, to arrange the information.

- CD layer has been adjusted to work on the nearby design of pixel information circulation, where a refined inclination drop improvement technique is utilized.

Class decomposition

The objective of our super sample decomposition component is to identify and leverage pseudo labels in the pretext training process of 4S-DT, using a collection of unlabelled images denoted as Y = {y_1, y_2.,y_n}. For this purpose, an autoencoder (AE) is initially employed to extract profound characteristics linked to every image. The representation vector h^d in equation (1) and reconstructed image *y*ˆin equation (2) can be defined for each input image *y* as.

$$h^d = f(W^{(1)}y + b^{(1)}) \tag{1}$$

$$\hat{y} = f(W^{(2)}h^d + b^{(2)}) \tag{2}$$

Where *f* ->Active Function; W -> Weight Matrices; b -> Bias Vector. The reconstruction error between input and reconstructed image is defined as shown in equation (3)

$$F(y,\hat{y}) = \frac{1}{2}\| y - \hat{y}^2 \| \tag{3}$$

The overall cost function of the n' unlabelled images, $E_{AE}(W,b)$, can be defined as shown in equation 4.

$$E_{AE}(W,b) = \left[\frac{1}{n^{'}} \sum_{i=1}^{n^{'}} L\left(y^{i}, \hat{y}^{i}\right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_i - 1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_i + 1} \left(W_{ji}^{(l)}\right)^2 \tag{4}$$

The initial term represents the reconstruction error of the complete dataset, while the subsequent term is the weight penalty term for regularization purposes. The latter term is intended to limit the weight magnitudes and prevent overfitting. The aforementioned variables are integral components of a neural network. Specifically, λ represents the weight decay

Table 4: Satellite data and its details

S. No.	Classes	Date of acquisition	Spectral Resolution	Spatial resolution
1	IRSP-6 (Resourcesat1) Multi-spectral	15-11-2018	Green (0.52-0.59 μm); Red (0.62-0.68 μm); Infrared (0.77-0.86 μm)	5.8 m

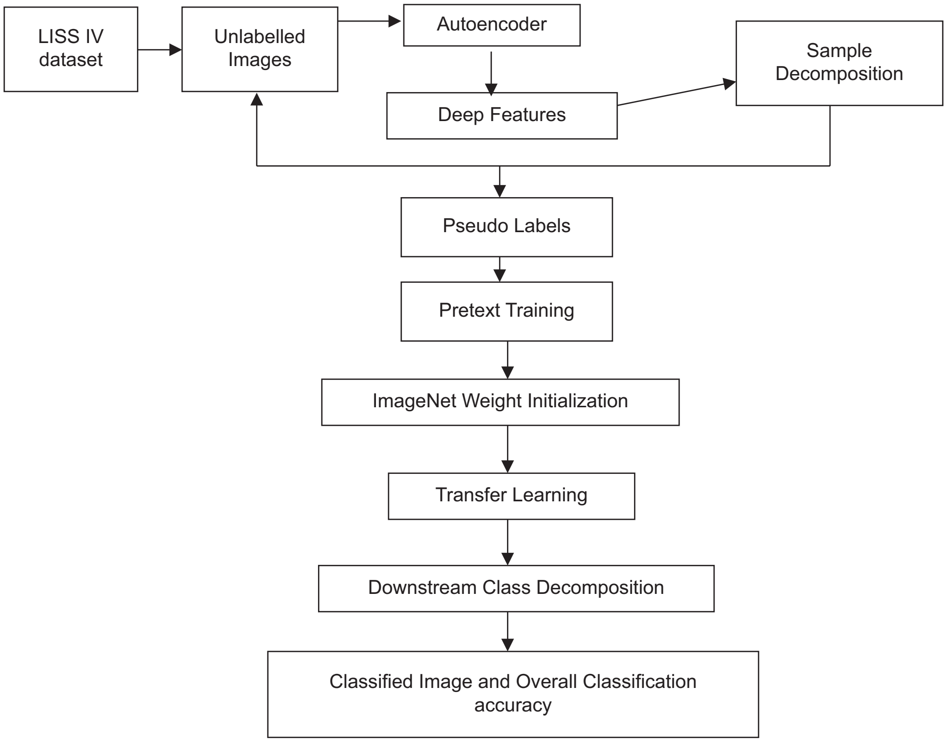


Figure 3: Architectural outline

parameter, n_l denotes the layer number of the network, s_l signifies the neuron number in layer l, and ${}_{i}W_{ji}^{(l)}$ refers to the connecting weight between neuron *i* in layer *l*+1 and neuron *j* in layer *l*.

Upon completion of the AE training, the image data distribution Y is subjected to Density-Based Spatial Clustering of Applications with Noise (DBSCAN) in order to classify it into multiple classes c, utilizing the extracted features hd. The DBSCAN algorithm is a type of unsupervised clustering method that is widely recognized as a significant density-based clustering approach. It characterizes clusters as the most extensive collection of points that are linked by density.

The given image dataset Y is projected onto a feature space of lowered dimensions, represented by $H \in R^{n'd}$. Here, H is a vector comprising of individual elements h_1, h_2., h_n. The density-based clustering algorithm considers two images y^i and y^j, represented by h_i and h_j respectively, to be density-connected in relation to *Eps* (i.e. neighbourhood radius) and *MinPts* (i.e. minimum number of objects within the neighbourhood radius of core object) if a core object y^k exists such that both y^i and y^j are directly density-reachable from y^k with respect to *Eps* and *MinPts*. In the context of density-based clustering, an image y^i is considered to be directly density-reachable from another image y^j if it falls within the Eps-neighbourhood of NEps(y^j) and y^j is classified as a core object as shown in equation 5. The Eps-neighbourhood is a mathematical construct and used to define a region around a given point, and is determined by a specified distance metric.

$$N_{Eps}\left(y_j\right) = \{y_i \in \mathrm{Y} \mid dis\ (y_i, y_{j)} \leq Eps\} \qquad (5)$$

The DBSCAN algorithm yields C clusters, wherein each cluster is formed by optimizing the density reachability correlation among images belonging to the same cluster. The n' unlabelled images will be assigned C cluster labels, which will serve as pseudo labels for both the pretext training task and the subsequent downstream training task. The dataset of pseudo-labelled images can be formally denoted as Y', where.Y''$=\{(y^i, x^c) \mid c \in C\}$.

Pretext training

Due to the widespread availability of extensively annotated image datasets, there is an increased probability that the various classes will be adequately represented. Thus, it is probable that the acquired knowledge within class boundaries is sufficiently generalizable to novel instances. Conversely, due to the restricted accessibility of annotated medical image data, particularly when certain categories are disproportionately underrepresented in terms of size and representation, there is a possibility of an escalation in the generalisation error. This phenomenon can occur due to a potential misalignment between the minority and majority categories. The utilization of extensive annotated image datasets, such as ImageNet, presents a viable approach to address this challenge through the process of transfer learning. This involves the training of CNN architectures, which necessitates the training of tens of millions of parameters.

During the adaptation and training process of an ImageNet pre-trained CNN model with a collected remote sensed dataset, a shallow-tuning mode was employed. The image feature space was constructed by utilizing pre-trained models on ImageNet, specifically the off-the-shelf CNN features, with training limited to the final classification layer.

The categorical cross entropy loss function, $E_{coarse}(\cdot)$, as shown in equation 6 was minimized using a mini-batch of stochastic gradient descent (mSGD).

$$E_{coarse}\left(x^{c},z^{'}\left(y^{j},W^{'}\right)\right)=-\sum_{c=1}^{C}x^{c}\ln z^{'}(y^{j},W^{'}) \qquad (6)$$

In this study, the set of self-labelled images in the training is denoted as y^{j}, while their associated self-labels are represented by x^{c}. The predicted output from a softmax function is denoted as $z^{'}(y^{j},W^{'})$, with the converged weight matrix associated to the ImageNet pre-trained model being represented by W'. It is worth noting that W' of the ImageNet pre-trained CNN model was utilized for weight initialisation to achieve a coarse transfer learning.

Downstream training

During the adaptation of the ResNet model, a fine-tuning mode was employed whereby feature maps from the coarse transfer learning stage were utilized. PCA was utilized to reduce the dimensionality of the images, as the high dimensionality of the feature space posed a challenge. This involved projecting the feature space into a lower dimension, where features that were highly correlated were disregarded. The aforementioned step holds significant importance in the subsequent phase of downstream training, as it facilitates the process of class-decomposition by producing classes that are more homogeneous. This, in turn, leads to a reduction in memory requirements and an improvement in the overall efficiency of the framework.

Let us consider the scenario where the feature space, obtained through PCA, is presented as a two-dimensional matrix, referred to as dataset A. In addition, let L denote a categorical variable representing the class. The symbols A and L can be expressed in an alternative manner as shown in equation 7.

$$A=\begin{bmatrix} a_{11} & a_{11} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}, \text{L=}\{ l_{1},l_{2}, \ \ldots\ldots\ldots,l_{c^{'}} \}(7)$$

where (n, m, c') stands for the total number of images, features, and categories. K-means clustering was employed for the purpose of downstream class decomposition (Löw *et al.*, 2013). This involved dividing each class into sub-classes that were homogeneous. The assignment of each pattern in the original class **L** to a class label was based on the nearest centroid, as determined by the squared Euclidean distance (SED) as shown in equation 8.

$$SED=\sum_{j=1}^{k}\sum_{i=1}^{n}\|\, a_{i}^{(j)}-c_{j} \qquad (8)$$

C_{j} is the centroid, when clustering is refined, the connection between dataset A and B can be numerically depicted as shown in equation 9

$$A=(A|L)\rightarrow B=B|C) \qquad (9)$$

Where the quantity of examples in An is equivalent to B while C encodes the new names of the subclasses (e.g., C'= $\{l_{11},l_{12}...,l_{1k},l_{21},l_{22}...,l_{2k}...,l_{c\,k}\}$).

Algorithm

With defined numerical definitions of the 2S-DT model, the procedural strides are exhibited and summed up in the algorithm.

Procedure

A. *Input*

- Remote sensed data is divided into training and testing sets.
- Ground Truth images with labels.

B. *Output*

- Classes are classified based on Labels which are predicted.

Self-supervised Decomposition

- Training an AE model to separate profound nearby highlights from the unlabelled Images.
- Apply an unaided grouping calculation for developing the pseudo-names.

Pretext Training

- Utilize an ImageNet pre-prepared by CNN model (for example ResNet18) for grouping of pseudo-named image dataset.
- Tweak boundaries are used on pre-training CNN model

Down Stream Task

A. *Class decomposition*

- Utilize an ImageNet pre-prepared CNN model (for example Alex Net) as an element extractor to remove highlights from info images.
- PCA is used for profound component space aspect.

- Utilize decreased element space of the info images to decay unique classes into various sub (or deteriorated) classes.

B. Fine exchange learning

- Adjust the last grouping layer in the CNN model to the deteriorated classes.
- Adjust boundaries of the affection preparing CNN model.

C. Class composition

- Compute the predicted labels and classify related to decomposed classes.
- Classified output

End Procedure

Parameter setting

The local size is a significant boundary that affects the characterization execution. To investigate the impact of neighborhood size on characterization precision, the area size is set to be 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, and 35, separately. Concerning Dataset 1, setting the area size to 26 or 28 has a small impact on the last grouping results. For Dataset 2, the neighborhood sizes of the data sets are set to be 27 × 27. AE was prepared, consisting of 78 neurons as the main hidden layer and 40 neurons as 2nd second hidden layer for the reconstruction of info un-labeled images. We utilized ResNet 18 (Li *et al.*, 2017) pretrained network to separate discriminative highlights of the marked Dataset.

- Input image size for Dataset 1 consists of 42*776 pixels and dataset 2 consist of 422*1056 pixels. A 3x3 kernel size was effective for capturing local features and patterns in the image while maintaining computational efficiency.
- Our ResNet architecture is structured with a series of residual blocks, with each block comprising two 3x3 convolutional layers. Following each convolutional layer, we incorporate batch normalization and apply a Rectified Linear Unit (ReLU) activation function.
- In our experiments, for layer Conv1 in ResNet18, we employed a 7x7 convolutional layer with 64 filters and a stride of 2. This configuration resulted in an output size of 112x112x64.
- For layer Conv2 in ResNet18, which includes Res2a and Res2b, the output size was set to 48x48x64.
- For Conv3 in ResNet18, encompassing Res3a and Res3b, we achieved an output size of 28x28x128. This architectural configuration was adopted for our experiments.
- Our ResNet engineering comprises lingering blocks, and each square has two 3×3 Conv layers, where each layer is trailed by cluster standardization and an amended straight unit (ReLU) initiation. Our ResNet design comprises lingering blocks, and each square has two 3×3 Conv layers, where each layer is trailed by clump standardization and a ReLU initiation work.
- During the training of the backbone network, we maintained a fixed learning rate of 0.0001 for all the layers, with the exception of the last fully connected layer, which had a learning rate of 0.01 to expedite learning.
- Our training utilized a mini-batch size of 256 samples and was conducted over a minimum of 200 epochs. To prevent overfitting during model training, we applied a weight decay of 0.0001 and set the momentum value to 0.95.
- Additionally, we employed a learning rate drop schedule, reducing it by a factor of 0.95 every five epochs. The training process for 2S-DT was executed in both shallow and fine-tuning modes.
- 4085 traits were set at this stage and utilized PCA to reduce the component elements. For the CD advance, we utilized k-implies bunching [8], where k is set to 2, and subsequently, each class in L has been additionally separated into two subclasses, resulting in a new Dataset with six classes.

For validating the proposed algorithm, the 3DCAE, TSVM, and GAN methods are used as benchmarks for classification. Detailed descriptions and parameters are provided as follows:

- 3DCAE (Mei *et al.*, 2019) is a solo spatial-spectral element learning strategy in light of a three-dimensional convolutional auto encoder. It is exceptionally successful in extricating spatial-spectral highlights.
- TSVM (LeCun *et al.*, 2015; Li *et al.*, 2017) is a semi-supervised technique that could utilize unlabeled samples to further develop characterization exactness. It likewise utilizes a SVM classifier with an outspread premise work part. All unlabeled samples are utilized for training.
- GAN (Radford *et al.*, 2015) is an unsupervised feature-learning technique in light of a generative adversarial networks. PCA is utilized to lessen the HSI to three aspects. Then, at that point, a 2-D GAN is utilized to learn highlights. The local size is too set to be 28 × 28.

RESULTS AND DISCUSSION

The performance of the 2S-DT algorithm was investigated for datasets 1 and 2 and compared with TVSM, 3DCAE, and GAN. Classification results are analyzed through qualitative and quantitative analysis with overall accuracy.

Visual Analysis

We also evaluate the visualization outcomes using the features derived from the original spectral features, the features derived from the MSE loss function, and the features derived from the contrastive loss function. In which different colors represent the distribution of spatial features in different classes by the selected samples. By observing this spatial distribution dataset, 1 and 2 features are distributed significantly and can show better classification results for all the algorithms.

Basically, all the maps provided in Figures 4-7 provided better classification results. Red and white circles are used to highlight the correct and incorrect classification results in the image. It was also observed that some salt and pepper noise was also identified in all the algorithms, and unstable behaviors were also observed with different classifiers, such as, for example: In dataset 1, the 2S-DT algorithm showed a quite good result when compared with other algorithms. When compared to TVSM, the dataset's 2S-DT and 3DCAE provided decent results, and

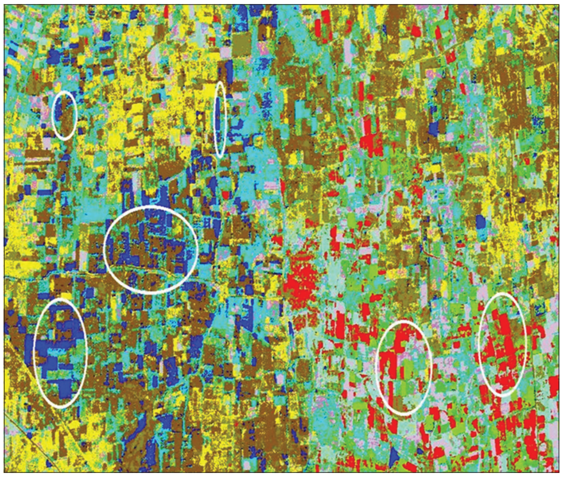


Figure 4: Image classified for six classes using TSVM (Dataset1)

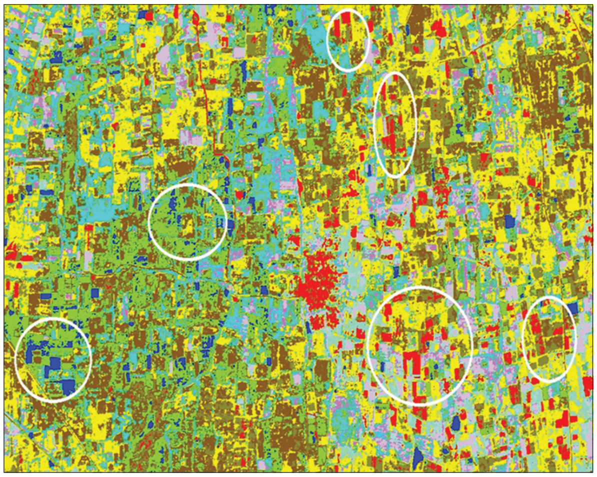


Figure 5: Image classified for six classes using 3DCAE (Dataset1)

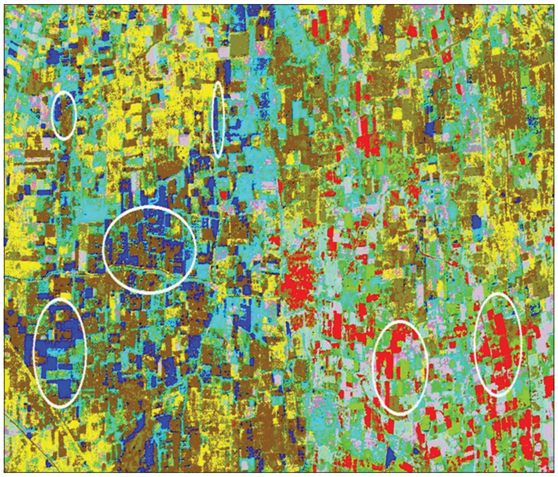


Figure 6: Image classified for six classes using GAN (Dataset 1)

the GAN algorithm produced slightly worse results. This may be due to the large number of mixed pixels in the datasets.

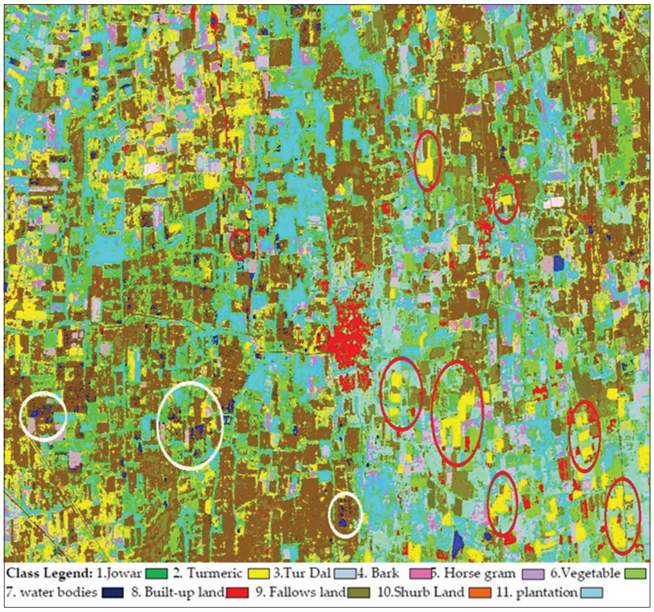

Class Legend: 1.Jowar 🟩 2. Turmeric 🟨 3.Tur Dal ▭ 4. Bark ▭ 5. Horse gram ▭ 6.Vegetable 🟩
7. water bodies ▬ 8. Built-up land 🟥 9. Fallows land ▬ 10.Shurb Land 🟧 11. plantation ▭

Figure 7: Image classified for six classes using 2S-DT (Dataset 2)

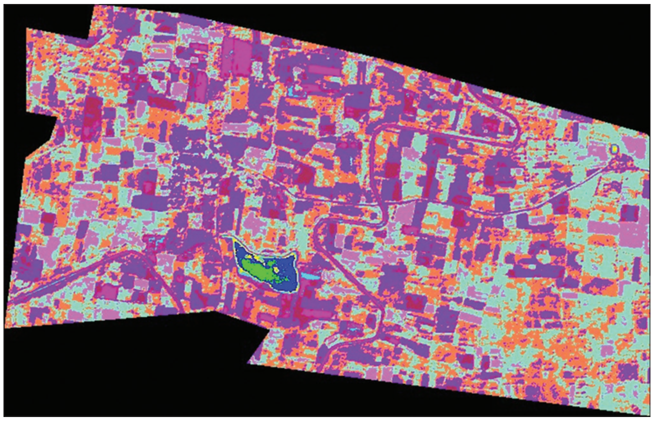


Figure 8: Image classified for twelve classes using TVSM (Dataset 2)

In addition to this, Figures 8-11 show that in TVSM and GAN algorithms, fallow land and shrub land were misclassified due to the influence of illumination conditions. Also, built-up land is misclassified with fallow land in 3DCAE due to the huge spectral similarity between these two classes. As the study area considered is rural areas, there is a lot of confusion between classes like others and annual crops. These problems are gradually reduced in the 2S-DT algorithm. At the same time, visual analysis of the 2S-DT algorithm shows reasonably smooth and acceptable results for fallow land when compared with other algorithms. In addition to this fallow land, Shurb land is classified correctly in the 2S-DT algorithm. For 3DCAE, built-up land is misclassified with fallow land in TVSM due to the huge spectral similarity between these two classes, which was classified correctly in the 2S-DT algorithm. Classes like Tur Dal and Horsegram were gradually reduced in the 2S-DT algorithm when compared with other algorithms. Besides, the misclassification of shrubs as vegetables was rectified and accurately identified in the 2S-DT algorithm.

Qualitative Analysis

To further evaluate the land cover performance of the 2S-DT method, classification results are compared with SVM and ABC.

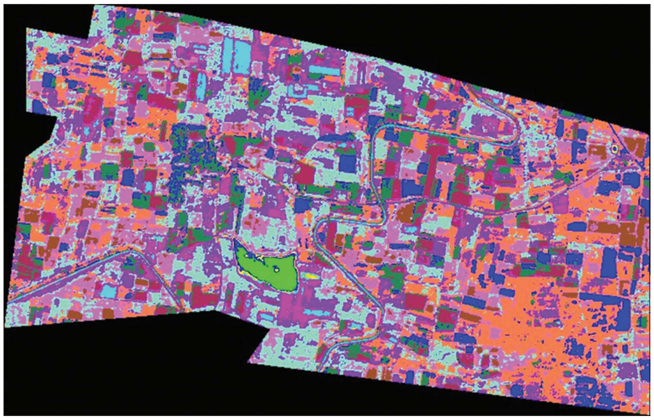


Figure 9: Image classified for twelve classes using 3DCAE (Dataset2)

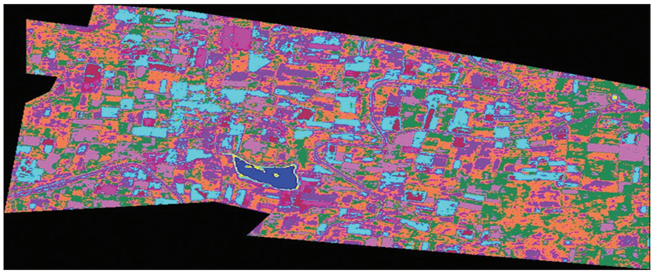


Figure 10: Image classified for twelve classes using GAN (Dataset 2)

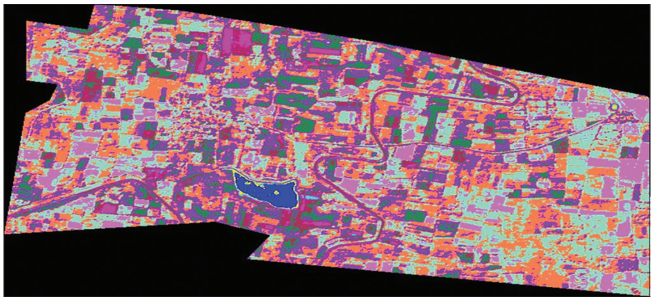


Figure 11: Image classified for twelve classes using 2S-DT (Dataset 3)

The 2S-DT algorithm was able to achieve the highest overall classification accuracy with all of the datasets. The performance of the annual crop is better in the 2S-DT algorithm when compared with another algorithm.

For dataset set 1, 2S-DT algorithms show an improvement when compared with 3DCAE, GAN, and TVSM. The classification results are tabulated in Tables 3-7. For the class build-up lands, 2S-DT shows a marginal improvement of 5% in PA when compared to 3DCAE and GAN but doesn't show any improvement in UA. For class fallow land, the 2S-DT algorithms indicate an improvement of 8.33% in PA when compared with GAN and 26.33% in PA when compared with 3DCAE. For class 5, the 2S-DT algorithm shows a marginal improvement when compared with GAN and TVSM. For class 6 (water bodies), 2S-DT and GAN show 100% in UA and 95.45% in PA, but it's marginally high when compared with 3DCAE.

Table 5: Confusion matrix for 3DCAE algorithm for Dataset1

Classes	1	2	3	4	5	6	Row Total	UA%
1	460			31	8	20	519	88.63
2		18					18	100
3			17		7	10	34	50.00
4	50			100	3	6	159	62.89
5	30			10	22		62	35.48
6	11		2	10	3	30	56	53.57
Column Total	551	18	19	151	43	66	875	
PA %	83.48	100	89.47	66.22	51.16	45.45		OCA73.94
Kappa	0.89	0.93	0.85	0.42	0.53	0.72		

1-AnnualCrop, 2-Waterbodies, 3-Built-upland, 4-Fallowland, 5-NaturalSpace and 6-others

Table 6: Confusion matrix for GAN algorithm for Dataset1

Classes	1	2	3	4	5	6	Row Total	UA%
1	530	1		10	1	1	543	97.60
2	5	16		20	2	1	44	36.36
3			17		1		18	94.44
4	15	1	2	140		1	158	88.60
5	1			8	39		48	81.25
6						63	63	100
Column Total	551	18	19	168	43	66	875	
PA %	96.18	88.88	89.47	83.33	90.69	95.45		OCA92.00
Kappa	0.92	0.84	0.86	0.82	0.86	0.90		

1-AnnualCrop, 2-Waterbodies, 3-Built-upland, 4-Fallowland, 5-NaturalSpace and 6-others

Table 7: Confusion matrix for proposed algorithm for Dataset 1

Classes	1	2	3	4	5	6	Row Total	UA%
1	542	1		4	1	1	549	98.72
2	2	17				1	20	85
3			19	10			29	65.53
4	7			154		1	162	95.06
5					42		42	100
6						63	63	100
Column Total	551	18	19	168	43	66	875	
PA %	98.36	94.44	100	91.66	97.67	95.45		OCA95.65

1-AnnualCrop, 2-Waterbodies, 3-Built-upland, 4-Fallowland, 5-NaturalSpace and 6-others

The study was extended to compare the performance of the 2S-DT algorithm in Dataset 2. The classification results are tabulated in Tables 8-11. The highest overall classification accuracy (OCA) of 88.91% was obtained in the 2S-DT algorithm when compared with GAN (OCA of 80.68%), 3DCAE (OCA of 62.63%), and TVSM (OCA of 59.08%). For the classes Water and Build-up Land, PA and UA show an improvement for the entire algorithm since they are spread less in this area and they are spatially homogenous. The UA of the plantation class remains the same for all the classifiers, but PA has an improvement of up to 6% when compared with 3DCAE and shows a marginal improvement with the GAN algorithm. For the class of land with or without shrubs, 2S-DT doesn't show marginal improvement when compared with GAN and 3DCAE. The class of fallow land is spatially homogeneous, it is not spectrally distinct. Hence,

Table 8: Confusion matrix of TVSM algorithm for Dataset 2

Classes	1	2	3	4	5	6	7	8	9	10	11	Row Total	UA %
1	127	1	3	22					15		3	171	74.26
2	13	36										49	73.46
3	43	2	50	9								104	48.07
4	52	1	4	84	1							142	59.15
5			5		33							38	86.84
6	1		1	23		15						40	37.5
7				9			18					27	66.66
8	1							19	101	1		122	15.57
9		2	5			4			62	4	16	93	66.66
10	2	2				2				38	12	56	67.85
11											35	35	100
Column Total	239	42	68	147	34	21	18	19	178	43	66	875	OCA59.08
PA %	53.36	85.71	73.52	57.14	97.05	71.42	100	100	34.83	88.37	53.03		
Kappa	0.63	0.72	0.35	0.42	0.63	0.72	0.68	0.69	0.43	0.58	0.67		

1-Jowar, 2-Turmeric, 3-TurDal, 4-Bark, 5-Horsegram, 6-Vegetable, 7-Waterbodies, 8-Built-upland, 9-Fallowsland, 10-ShurbLand and 11-Plantation

Table 9: Confusion matrix for 3DCAE algorithm for Dataset2

Classes	1	2	3	4	5	6	7	8	9	10	11	Row Total	UA %
1	127			12					101			240	52.91
2	7	39										46	84.78
3	2	2	58						14			76	76.31
4	53			84								137	61.31
5	1		3		33				1		5	43	76.74
6	4			9	1	15					4	33	45.45
7							18			4		22	81.81
8			7					19				26	73.07
9		1		10		4			62			77	80.51
10				32						38	2	72	52.77
11	32					2				1	55	58	94.82
Column Total	239	42	68	147	34	21	18	19	178	43	66	875	OCA62.63
PA %	53.14	92.85	85.29	57.14	67	71.42	100	100	34.83	88.37	83.33		
Kappa	0.64	0.56	0.43	0.60	0.57	0.87	1	1	0.79	0.23	0.72		

1-Jowar, 2-Turmeric, 3-TurDal, 4-Bark, 5-Horsegram, 6-Vegetable, 7-Waterbodies, 8-Built-upland, 9-Fallowsland, 10-ShurbLand and 11-Plantation

Table 10: Confusion matrix for GAN algorithm for Dataset2

Classes	1	2	3	4	5	6	7	8	9	10	11	RowTotal	UA %
1	169		7	41								217	77.88
2	10	38	6									54	70.37
3	2	1	54									57	94.73
4	37			102		5			10			154	66.23
5	16	1			33							50	66
6	3	2		2	1	15						23	65.21
7			1				18					19	94.73
8								19	42		7	68	61.76
9	2			1					169	11	1	184	91.84
10				1		4			16	32	1	54	59.29
11									2		57	59	96.61
Column Total	239	42	68	147	34	21	18	19	178	43	66	875	OCA80.68
PA %	70.71	90.47	79.41	69.38	97.05	71.42	100	100	94.94	74.41	86.36		
Kappa	0.67	0.59	0.74	0.61	0.79	0.98	1.0	1.0	0.82	0.42	0.88		

1-Jowar, 2-Turmeric, 3-TurDal, 4-Bark, 5-Horsegram, 6-Vegetable, 7-Waterbodies, 8-Built-upland, 9-Fallowsland, 10-ShurbLand and 11-Plantation

it shows a marginal improvement in 2S-DT when compared with GAN (4% in PA and 2% in UA) and 3DCAE (3% in PA and 2% in UA). This also results in an improvement in the Kappa value from 0.63 to 0.881 in comparison to other algorithms. As Jowar is one of the major crops grown in the winter region during this season, this class shows an improvement of 10% in PA and 5% in UA when compared with other algorithms. For class Tur Dal, most of the misclassification has come from the BARK crop, and a major portion has also been misclassified. The GAN algorithm reveals an improvement in PA and UA

Table 11: Confusion matrix for proposed algorithm for Dataset 2

Classes	1	2	3	4	5	6	7	8	9	10	11	RowTotal	UA %
1	214		6	10	1						1	232	92.24
2		40									2	42	95.23
3	13	2	56	2								83	67.46
4	10		6	130	1						5	142	84.50
5	2			3	32							37	86.48
6				2		15						17	88.23
7							18					18	100
8								19	8			27	70.37
9						2			170	17		189	89.94
10						4				26		30	86.66
11											58	58	100
Column Total	239	42	68	147	34	21	18	19	178	43	66	875	OCA88.91
PA %	89.53	95.23	82.35	81.63	94.11	71.42	100	100	88.76	60.46	87.87		
Kappa	0.85	0.92	0.83	0.79	0.92	0.78	1.0	0.92	0.88	0.75	0.89		

1-Jowar, 2-Turmeric, 3-TurDal, 4-Bark, 5-Horsegram, 6-Vegetable, 7-Waterbodies, 8-Built-upland, 9-Fallowsland, 10-ShurbLand and 11-Plantation

Table 12: Execution Time of training and feature extraction

	TVSM	3DCAE	GAN	Proposed method
Dataset1				
Training (Min)	20.36	19.23	6.54	32.23
Feature Extraction (Sec)	31.26	15.32	3.56	20.23
Dataset2				
Training (Min)	19.32	19.37	7.02	82.00
Feature Extraction (Sec)	5.22	32.04	0.54	40.13

when compared with 2S-DT and 3DCAE. For the spectrally overlapping classes such as turmeric and vegetables, 2S-DT has shown a PA of 83.33% and a UA of 85.26% for both classes. On the contrary, GAN and 3DCAE have shown a fall of 5% on PA and UA for turmeric class and 7% on vegetables. For class Horesgram, the 2S-DT algorithm exhibits a PA of 100% and a UA of 93.75%, which is high when compared with other algorithms. There is no serious misclassification among other classes. For class Bark 2S-DT algorithm, the performance of all the algorithms is almost the same.

Execution Time

Input dimensions, network parameters, and the quantity of samples used during TP all affect the training period (TP) of a deep neural network. The 2S-DT model needs less time when compared with other algorithms when the number of classes is under Dataset 1 and requires more time when the number of classes is under Dataset 2. As the 2S-DT approach is a learning based algorithm, different hierarchy levels may lead to different feature extraction times. Table 9 shows the comparison of different methods with the 2S-DT method. The 2S-DT method takes about 32.23 minutes to train the data set and 20.23 seconds to extract the features at Dataset 1, 82 minutes to train the data set, and 40.13 seconds to extract the features at Dataset 2. However, there is an increase in execution time and training time when compared with other methods, but the 2S-DT method shows an improvement in classification accuracy in all the hierarchy levels and datasets (Table 12). All the studies are performed on a CPU using Python software. As can be observed, the 2S-DT algorithm is the fastest for the selected study area when compared to other algorithms.

CONCLUSIONS

The 2S-DT (Self-Supervised Decomposition for Transfer Learning) model, a ground-breaking approach to crop categorization utilizing high-resolution remote sensing data, is introduced as a result of our study. This methodology successfully handles the persistent problem of misclassification, particularly when dealing with unlabeled classes. To maximize information structure depending on geographical context, the 2S-DT model makes use of self-supervised learning approaches and adds a Class Decomposition (CD) layer. Our model uses residual blocks with two 3x3 convolutional layers, followed by batch normalization and ReLU activation functions, and is based on the reliable ResNet architecture. The output size of ResNet18 is 112x112x64 because to our careful architecture decisions, which include using a 7x7 convolutional layer with a stride of 2 for Conv1. The output size of Conv2, which includes Res2a and Res2b, is 48x48x64, while the output size of Conv3, which includes Res3a and Res3b, is 28x28x128. These choices are made in accordance with the details of our experiments. The usefulness of the 2S-DT model has been demonstrated through extensive experimentation on two datasets. It achieves 95.65% accuracy for dataset 1 and 88.91% accuracy for dataset 2, showing a considerable improvement in overall accuracy. In particular, our model performs better than counterpart models like TVSM, 3DCAE, and GAN Model. We also emphasize the advantages of the 2S-DT method in our comparison study. Dataset 1 demonstrates advancements in a number of significant classifications, such as "build-up lands" and "fallow land." The model outperforms rival algorithms in Dataset 2 by achieving the greatest overall classification accuracy (OCA) of 88.91%.

Our findings pave the way for more investigation and improvement in the future. The continuing improvement of the 2S-DT model is one path that shows promise. Further improvements in classification performance may be unlocked by fine-tuning hyperparameters and investigating alternate setups. It's intriguing to think about using the 2S-DT algorithm for more diverse remote sensing jobs. In order to determine its adaptability to various geographic and environmental situations, further research will be needed. Additionally, given the model's efficacy in differentiating spectrally overlapping classes like "turmeric" and "vegetables," additional study may explore related problems in a variety of areas.

ACKNOWLEDGMENT

The authors would like to thank NRSC for providing the data and GSSSIETW and Navkis College for software support throughout this work. This research was funded by "Consolidated University Research for Innovation and Excellence CORE Grant for Women PG College" (DST-CURIE), Department of Science and Technology, Govt. of India, New Delhi, India, "DST/CURIE-PG/2022/71".

REFERENCES

Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Research frontier: deep machine learning - A new frontier in artificial intelligence research. *IEEE Computational Intelligence Magazine, 5*(4), 13-18. https://doi.org/10.1109/MCI.2010.938364

Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology, 173*, 74-84. https://doi.org/10.1016/j.agrformet.2013.01.007

Bruzzone, L., Chi, M., & Marconcini, M. (2005, July 29). Transductive SVMs for semisupervised classification of hyperspectral data. Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. (pp. 4). IEEE. https://doi.org/10.1109/IGARSS.2005.1526130

Esch, T., Metz, A., Marconcini, M., & Keil, M. (2014). Combined use of multi-seasonal high and medium resolution satellite imagery for parcel-related mapping of cropland and grassland. *International Journal of Applied Earth Observation and Geoinformation, 28*, 230-237. https://doi.org/10.1016/j.jag.2013.12.007

Gallego, J., Kravchenko, A. N., Kussul, N. N., Skakun, S. V., Shelestov, A. Y., & Grypych, Y. A. (2012). Efficiency assessment of different approaches to crop classification based on satellite and ground observations. *Journal of Automation and Information Sciences, 44*(5), 67-80. https://doi.org/10.1615/JAutomatInfScien.v44.i5.70

Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment, 188*, 9-25. https://doi.org/10.1016/j.rse.2016.11.004

Hedhli, I., Moser, G., Zerubia, J., & Serpico, S. B. (2016). A new cascade model for the hierarchical joint classification of multitemporal and multiresolution remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing, 54*(11), 6333-6348. https://doi.org/10.1109/TGRS.2016.2580321

Jayanth, J., Shalini, V. S., Kumar, T. A., & Koliwad, S. (2020). Classification of field-level crop types with a time series satellite data using deep neural network. In D. Hemanth (Eds.), *Artificial Intelligence Techniques for Satellite Image Analysis* (Vol. 24, pp. 49-67) Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-24178-0_3

Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters, 14*(5), 778-782. https://doi.org/10.1109/LGRS.2017.2681128

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444. https://doi.org/10.1038/nature14539

Li, Y., Zhang, H., & Shen, Q. (2017). Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing, 9*(1), 67. https://doi.org/10.3390/rs9010067

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS Journal of Photogrammetry and Remote Sensing, 85*, 102-119. https://doi.org/10.1016/j.isprsjprs.2013.08.007

Mathur, A., & Foody, G. M. (2008). Crop classification by support vector machine with intelligently selected training data for an operational application. *International Journal of Remote Sensing, 29*(8), 2227-2240. https://doi.org/10.1080/01431160701395203

Mei, S., Ji, J., Geng, Y., Zhang, Z., Li, X., & Du, Q. (2019). Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing, 57*(9), 6808-6820. https://doi.org/10.1109/TGRS.2019.2908756

Omkar, S. N., Senthilnath, J., Mudigere, D., & Kumar, M. M. (2008). Crop classification using biologically-inspired techniques with high resolution satellite image. *Journal of the Indian Society of Remote Sensing, 36*, 175-182. https://doi.org/10.1007/s12524-008-0018-y

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint. arXiv:1511.06434. https://doi.org/10.48550/arXiv.1511.06434