Journal of King Saud University – Computer and Information Sciences xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

CaneSat dataset to leverage convolutional neural networks for sugarcane classification from Sentinel-2

Shyamal S. Virnodkar[a,*], Vinod K. Pachghare[a], V.C. Patil[b], Sunil Kumar Jha[b]

[a] Department of Computer Engineering & IT, College of Engineering Pune, Savitribai Phule Pune University, India
[b] K J Somaiya Institute of Applied Agricultural Research (KIAAR), Sameerwadi, Mudhol Taluka, Bagalkot District, 587316, India

ARTICLE INFO

Article history:
Received 30 April 2020
Revised 3 August 2020
Accepted 8 September 2020
Available online xxxx

Keywords:
Multi-temporal remote sensing images
Sugarcane classification
Convolutional neural network
Deep network

ABSTRACT

The ubiquitous deep learning (DL) in remote sensing (RS) motivates the most challenging problem of crop classification. To perpetrate such an exigent task, an attempt is made to prepare a novel dataset, the CaneSat dataset, in two formats: RGB color space and geo-tiff images, covering the region of four talukas in Karnataka, India. This research aims to build a model for sugarcane classification using two-dimensional convolutional neural network (CNN or ConvNet) applying RS time series data. Further, the study intents to evaluate competency of four state-of-the-art deep CNNs namely AlexNet, GoogLeNet, ResNet50 and DenseNet201 using fine tuning and deep CNNs as feature extractors to classify sugarcane and non-sugarcane areas from Sentinel-2 data. The results of the research are expressive on CaneSat dataset. It shows that the CNN model performs significantly good producing 88.46% accuracy, whereas all deep networks exhibit more than 73.00% overall accuracy. When used as feature extractors, ResNet50 and DenseNet201 outperform all other models with precision of 85.65% and 87.70%, respectively. Noticeably, the results indicate that 2D CNN model and features extracted using CNNs with SVM classifier are efficient methods for sugarcane classification from Sentinel-2 time series data in peninsular zone of India.

1. Introduction

Agriculture accounts for over 50% of Indian population livelihood and is backbone of the Indian economy and food system. Sugarcane is a cash crop of India and sugar mills wants to know the cane availability, so that they can plan their harvesting schedule. Field assistants are assigned this job to get the information of cane availability which eventually leads to human error and mills shortfall in crushing every year. Sugarcane classification at every stage would not only help mills but also the farmers who are holding significantly large area to manage their farm. Earth observation (EO) becomes powerful technology to achieve this challenging task. EO provides continuous, autonomous, high quality dataset with a global coverage of earth observation. With open access to such a huge amount of satellite data, abundant applications in the domains of agriculture (Virnodkar et al., 2019a, 2020) and urban development (Ponti et al., 2016) have successfully been realized. The high temporal revisit period becomes a powerful source for time series datasets that can be useful for monitoring geographical area and vegetation dynamics (Zheng et al., 2020) through time. How to analyze and utilize this time series to leverage the seasonal characteristics of vegetations varying with time and season is still an unfastened issue in the RS research field. Notwithstanding the usefulness of the time series, the traditional approach is to execute futuristic ML techniques like SVM and RF on stacked satellite images (Yang et al., 2011). Time series data exhibits temporal correlations which are failed to model by these traditional approaches, as they extract the features autonomously from one another regardless of temporal dependencies.

Recently, DL technology have achieved astonishing performance in crop and land use land cover (LULC) classification from RS time series images, in particular; the CNN and the long short-term memory (LSTM) which is a gated recurrent unit of recurrent neural network (RNN). Prior to the development of DL, the RS

* Corresponding author.
E-mail addresses: ssv18.comp@coep.ac.in (S.S. Virnodkar), vkp.comp@coep.ac.in (V.K. Pachghare), patil.vc@somaiya.com (V.C. Patil), jha.sunilkumar@somaiya.com, umar@somaiya.com (S.K. Jha).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2020.09.005

community have focused on the employment of SVM and ensemble classifiers for various RS applications in particular crop classification (Belgiu and Csillik, 2018; Virnodkar et al., 2019b) from the use of a neural network (Atkinson and Tatnall, 1997) which was the basis of DL algorithms (Ma et al., 2019). In this research, an attempt is made to classify sugarcane and non– sugarcane using CNN from Sentinel-2 time series data in the region of Mudhol, Jamkhandi, Raibag and Gokak talukas of Karnataka, India. It is a 6 layers network and achieves an accuracy of 88.46%. To leverage the pre-trained DL models, we have also evaluated four models in this study for the target dataset of sugarcane.

In this paper, following contributions are presented

1. A dataset ('CaneSat'), containing total of 1627 sugarcane and non-sugarcane images, is created. The dataset is in two formats: one contains geo-tiff images with six features and another having jpg images in RGB colour space. Geo-tiff images are geo-referenced and labeled, however, jpg images are not geo-referenced.
2. A model using DL is proposed on the created dataset to learn spatial features from RS time series data for classification.
3. Comparative performance evaluation of AlexNet, GoogLeNet, ResNet50 and DenseNet201 models' transfer learning has been carried out on the created 'CaneSat' dataset.
4. RF and SVM ML techniques have also been tested to compare CNNs' performances.

1.1. Related work

The advancement of DL in the field of RS increases with the continuous generation of a huge data by satellites. Nonetheless, big labeled data in RS is rare and is a very time consuming and labour and cost intensive process, however, is an essential requirement for DL frameworks. Since 2010, many researchers put their efforts and produced small as well as large scale RS labeled datasets. The most studied and sought-after dataset is the UC Merced (UCM) dataset created from aerial images by Yang and Newsam (2010) for land use classification. Novel datasets generated from the Google earth images having high SR are NWPU-RESISC45 (Cheng et al., 2017) and PatternNet (Zhou et al., 2018). Dataset for object detection in aerial images (DOTA) (Xia et al., 2018) and aerial image dataset (AID) (Xia et al., 2017) are two large scale datasets aid research in RS domain. Many other RS datasets include the Brazillian coffee scene (BCS) dataset (Penatti et al., 2015), the SAT-4 and the SAT-6 (Basu et al., 2015), the RS19 (Nogueira et al., 2017) and the EuroSat (Helber et al., 2019).

Several research studies gained significant results on the available hyper-spectral and multispectral RS datasets (Ienco et al., 2017; Interdonato et al., 2019), by training new CNNs from scratch. Another technique, fine tuning in transfer learning freezes initial layers and adjust parameters of the last layers according to the target dataset. Many studies reported state-of-the-art results of fine tuning on different RS datasets (Penatti et al., 2015; Mahdianpari et al., 2018). The pre-trained CNNs performed well as feature extractors (FE) to extract deep features from images (Castelluccio et al., 2015, Hu et al., 2015).

From the literature survey, it is observed that various datasets like UCM, BCD, NWPU- RESISC45, EuroSat, SAT-4, SAT-6, etc. are openly available, however, a few of them are not publicly available and none of them cover Indian geography. Most of the datasets are used for LULC and crop land mapping which contains sugarcane crop as one of the class.

However, due to environmental variability, every geographical location has shown to influence crop canopy structure, crop class, foliar chemistry which can change the satellite imagery reflectance value. Hence, there was a need to create a dataset to perform sugarcane classification in the study area located in India. In the rest of the paper, materials and methods are explained in Section 2; results are presented and discussed in Section 3 and finally concluded the paper in Section 4.

2. Materials and methods

The research study area as depicted in Fig. 1 spreads across four talukas, namely, Mudhol, Jamkhandi, Raibag and Gokak covering 8 lack acres of land in Karnataka, India at 16.38980° N and 75.03710° E. The area has an altitude of 541 m above sea level with annual precipitation is around 545 mm. The climate is generally dry and the temperature ranges between 16.20 °C and 38.70 °C. Sugarcane is the main crop cultivated in this region. Sugarcane is a semi-perennial and one of the most prime crops across the world, especially in India, Brazil and China. Brazil ranked first in sugar production and second in ethanol production. There are three plantation seasons in south India like early (Jan – Feb), mid-late (Oct – Nov) and late (Jul – Aug). It undergoes four growing phases, namely, germination phase, tillering phase, grand growth phase and maturity phase in south and north India.

2.1. Formation of the CaneSat dataset

Abundant applications of agriculture at a large scale can be enhanced with the use of freely available RS data mainly crop classification; motivates the formation of the proposed dataset. All satellite images utilized in this study are cloud free and downloaded from ESA's Copernicus program website (home). In the RS context preprocessing of the images are required in order to remove atmospheric effect, radiometric noise and geometric errors. This preprocessing normalizes the data. The available Sentinel-2 data is geometrically and radiometrically corrected. However, all the images have undergone atmospheric correction as the normalization step which was performed through SCP plugin tool in GIS 2.18. Eight images acquired during Oct 2018 to May 2019 are utilized to generate the dataset. The aspiration behind covering four talukas for preparing dataset is to train the network with the high variance intrinsic to the satellite images due to data capturing, pre-processing and other parameters affecting the sugarcane crop growth, like irrigation type, crop type, crop variety, soil texture, soil moisture, soil type, climate, precipitation, temperature and humidity. The sugarcane raising at different talukas varies little bit in their reflectance captured by the remotely sensed images due to the above-mentioned parameters. Hence, phenology of sugarcane from different talukas is covered in the CaneSat dataset. The parameters considered in collecting sugarcane samples is given in Table 1.

The CaneSat dataset formation process is depicted in Fig. 2. There are two classes in the dataset, one is sugarcane, comprises all growing phases and the non-sugarcane class consists of all other land covers existing in the study area. It includes maize, built-up, residential, industrial building, water body, pastures, rocks and fallow land. Different contextual window sizes of 28 × 28, 64 × 64 and 256 × 256 have been employed in the various studies to create the land use land cover dataset from the satellite images. However, in our study area, most of the farmers own small land (~1 acre or 0.404 ha) for agriculture or cultivate different crops adjacent to each resulting in a small patch of images in the dataset. The geo-referenced image patches of 10x10 pixels are clipped from the Sentinel-2 A/B satellites with each pixel of 10 m resolution. Ground truth data is a geographical information system (GIS) shape files collected through fields' survey conducted during October 2018 to May 2019. Ground truth data has been recorded by the global positioning system (GPS) device (Montana 680) for

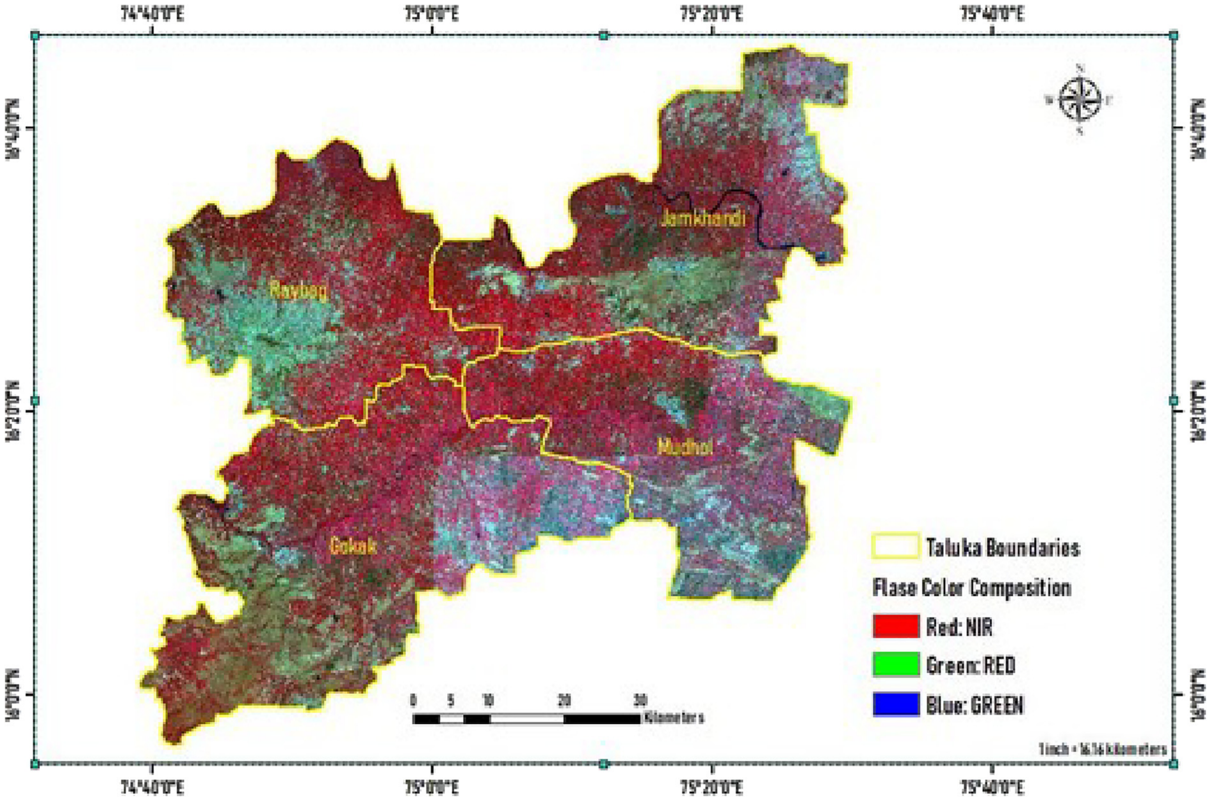


Fig. 1. The study area.

Table 1
Parameters of collected sugarcane samples.

Sugarcane Variety	CO265	VSI8005	SNK9293	CO91010	CO86032
Count	34	46	56	63	671
Soil type	Light Black	Mixed	Red	Medium Black	Black
Count	25	42	90	158	555
Plantation type	Ratoon 3	Ratoon 2	Ratoon 1	Plantation	
Count	10	30	299	531	
Talukas	Mudhol	Jamkhandi	Raibag	Gokak	
Count	260	200	200	210	
Irrigation type	Drip	Flood			
Count	192	678			

sugarcane and maize crops. Besides, samples of all other objects in the non-sugarcane class were generated through a visual interpretation based on experts' knowledge. All vector files of 10 × 10 are precisely drawn to assure geometry of the image patches as shown in Fig. 3.

Then, all GIS shape files are converted into raster files giving an image patch of 10 × 10 pixels. Altogether, the complete ground truth data contains 1627 samples having 162,700 pixels covering an area of 16.27 ha. Out of which 87,000 pixels are for sugarcane class and remaining 75,700 pixels for non-sugarcane class. The main contribution of this research study is that the dataset is emancipates in two formats i.e. jpg format in RGB colour space and tif format of geo-referenced images (Fig. 4), aiming to accelerate the use of machine learning in crop classification task using openly accessible RS data. The tif format is composed of six spectral bands such as Red, Green, Blue, Near Infrared (NIR), Red edge, Short-Wave Infrared (SWIR). In total, six features are examined for every pixel in every image of the time series of eight images. The CaneSat dataset is publicly available at 'https://ieee-dataport.org/documents/canesat'.

2.2. Convolutional neural networks

DL models learn features from the images automatically, unlike the state-of-the-art ML methods. In the last few years, DL is in vogue among researchers for processing RS data. Among all the DL models, CNNs are becoming increasingly ubiquitous by the reason of their remarkable results (Krizhevsky et al., 2012) in many domains including RS (Yan et al., 2019; Nogueira et al., 2015). This is because of the image's stationary property which states that contents retrieved from one part of an image can also be applied to other part of the image (Nogueira et al., 2015). CNN architecture comprises many layers, namely, convolutional layers (comprises processing units, i.e., neurons), sub-sampling layers and fully connected (FC) layers with nonlinear transformations, refers to the deep architecture. The convolutional layer acts as a function to extract features from previous layers of the network, generates feature maps as an output of every layer. The feature map values depend on the structure of the kernel which defines what information is to be extracted from the layers. The kernel comes in a matrix and is responsive to the spatial information of an image. Features are extracted at different levels as low, mid and high level from initial, middle and final layers, respectively. Initial layers are responsible for extracting more generic features like color blob, corners and edges which are not application oriented. Final layers are more specific to the application, extract the objects or image structures and therefore, need to be trained according to the application and target dataset. Generally, but not necessarily, each convolutional layer is followed by a sub-sampling layer. The sub-sampling layer, also called as pooling layer, is mainly implemented to reduce size of the image data and the variance of features extracted from the convolutional layer with the retention of the geometry of the input data (Rezaee et al., 2018). A fixed-size grid runs over the image feature map

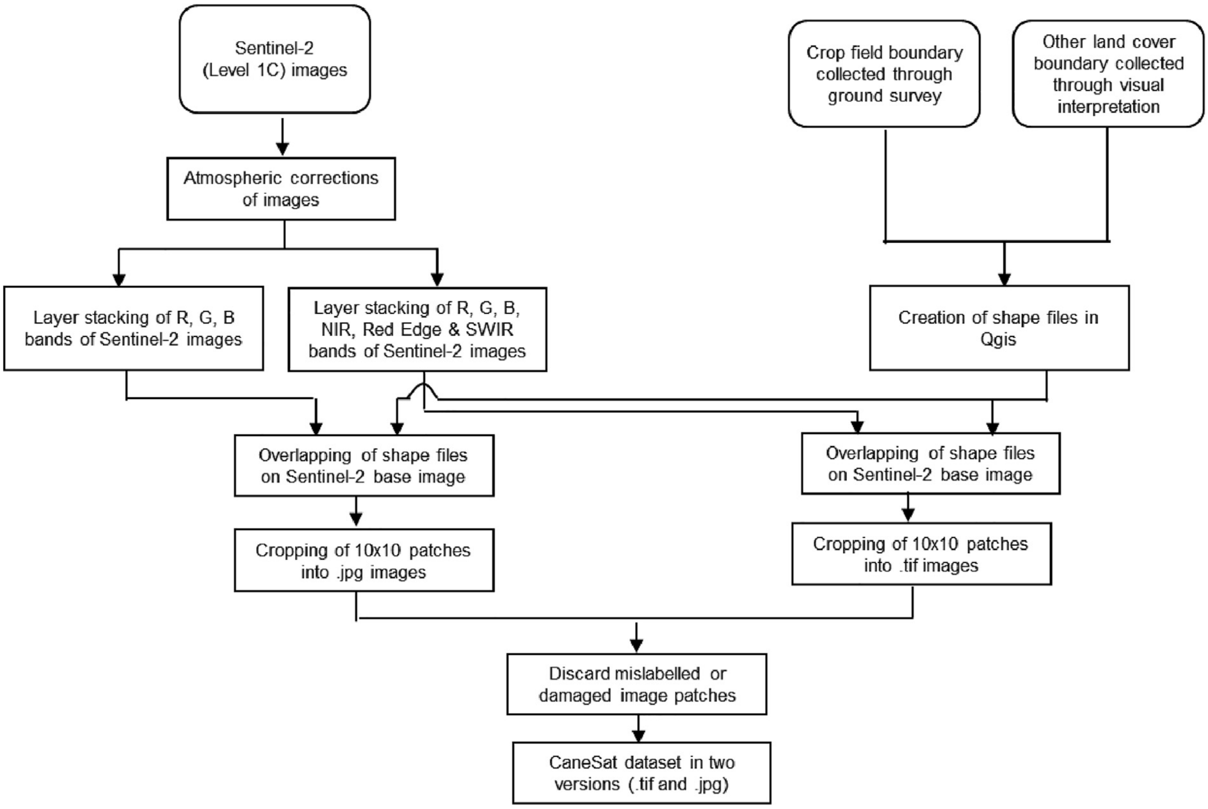

Sentinel-2 (Level 1C) images
Atmospheric corrections of images
Layer stacking of R, G, B bands of Sentinel-2 images
Layer stacking of R, G, B, NIR, Red Edge & SWIR bands of Sentinel-2 images
Crop field boundary collected through ground survey
Other land cover boundary collected through visual interpretation
Creation of shape files in Qgis
Overlapping of shape files on Sentinel-2 base image
Overlapping of shape files on Sentinel-2 base image
Cropping of 10x10 patches into .jpg images
Cropping of 10x10 patches into .tif images
Discard mislabelled or damaged image patches
CaneSat dataset in two versions (.tif and .jpg)

Fig. 2. CaneSat dataset formation process.

with a stride and implements max (or average) operation which gives maximum (or mean). The sub-sampling layers select the most important, robust and abstract features (Mahdianpari et al., 2018) for the following layers which aid in lowering the computational difficulty in the network training. Over many succeeding convolutional and pooling layers, FC layers exist till the classifier in the network architecture. There could be more than one FC layers, however, the last FC layer connects all processing units of the previous layer to its every single processing unit and concludes the output to the classifier layer. Lastly, the classifier layer is employed which computes the posterior probability of each class instance. A softmax function is a widely used classifier layer, also known as a normalized exponential, represents categorical probability distribution to predict the probability of the sample class (Nogueira et al., 2017). In addition to all layers, normalization layers, local response normalization (LRN) and batch normalization are generally used with unbounded activations such as rectified linear unit (ReLU) to detect high frequency features. The CNN was anteceded way back in the 1980s with the design of a primary CNN, LeNet, by LeCun et al. (1998) for handwritten digits classification. However, it was progressively employed in many domains since 2010. Thanks to the development of GPUs and larger datasets like ImageNet (Deng et al., 2009), Cifar (Krizhevsky, 2009) those contributing in advancing the network design. A quantum leap in the development of deep CNN was by Krizhevsky et al. (2012) with the design of AlexNet which becomes a foundation to modern deeper feature learning CNNs. Thereafter, since 2014 further manoeuvre has successfully been achieved with the evolvement of CaffeNet (Jia et al., 2014) inherited from AlexNet, VGG (Simonyan and Zisserman, 2014) and its variants, GoogLeNet (i.e., Inception network) (Szegedy et al., 2015), ResNet (He et al., 2016) and DenseNet (Huang et al., 2017).

2.2.1. Alexnet

AlexNet (Krizhevsky et al., 2012), secured first place in the ImageNet large scale visual recognition challenge (ILSVRC) held in 2012, for object detection. This network includes a total of eight hidden layers having five convolutional layers with three max pooling layers and three FC layers. The nonlinearity was added by ReLU function and LR normalization was implemented after first and second convolutional layers. The endmost FC layer is succeeded by a softmax activation layer. AlexNet was successful by dint of the realization of GPUs for convolution operations, use of dropout to overcome overfitting problem at FC layers, non-saturating neurons and more training samples. In addition to this, it requires fewer parameters, i.e., 60 million and 650,000 neurons which reduces network training time.

2.2.2. GoogLeNet

GoogLeNet was developed by (Szegedy et al., 2015), a deeper convolutional network also codenamed Inception-v1, won first place in the ILSVRC 2014 image classification contest with lower error rate compared to the VGGNet. GoogLeNet has total of 22 layers (plus five pooling layers) comprises 1 × 1 convolutional layer as the first layer, nine inception modules realizing network in network approach, softmax as auxiliary classifiers used for training only and a global average pooling layer at the end instead of FC layer. This CNN is advantageous over previous CNNs because of i) 1 × 1 convolutional layer lessens the model, ii) different convolutional types along with different max pooling for same input to retrieve more spatial details, ii) efficiently handled overfitting problem by turning down the number of parameters used in the network and employment of global average pooling layer iv) auxiliary classifiers hinders vanishing gradient problem in the network's deeper approach and provides regularization as well.

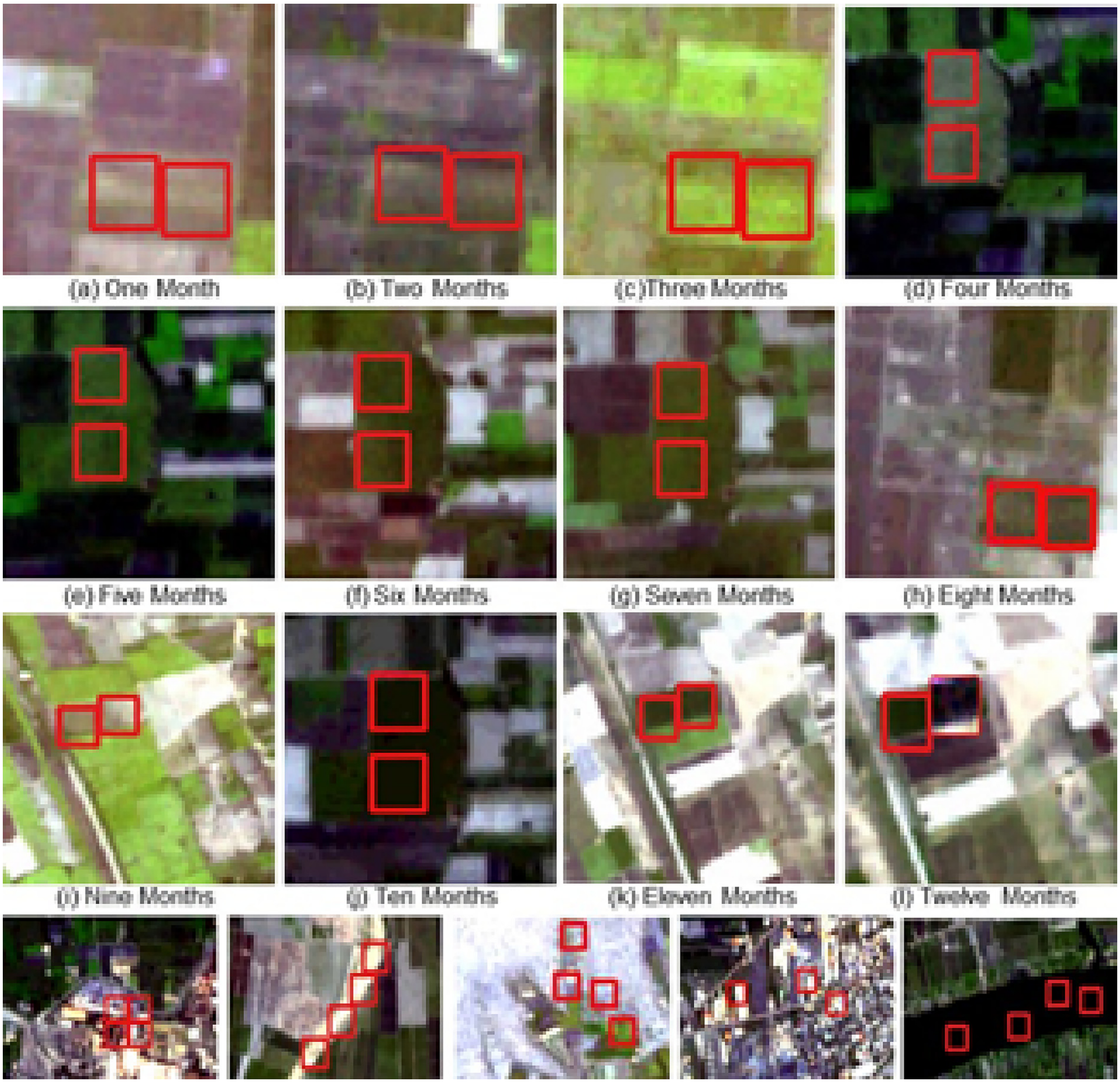


Fig. 3. Image samples bounding boxes of sugarcane class and land cover class.

2.2.3. ResNet50

ResNet, a residual neural network (He et al., 2016) was winner of the ILSVLC 2015 competition with 3.57% top-5 error rate. ResNet also secured 1st place in ILSVRC and COCO 2015 competition in ImageNet localization, ImageNet detection, Coco segmentation and Coco detection (He and Sun, 2015. In addition to the difficulty in training deep networks, they all were suffered from the major problem of vanishing gradient which makes learning infinitesimal as it is back propagated in the initial layers. Prior to ResNet, few deep architectures tried to get rid of vanishing gradient problem. However, ResNet succeeded by introducing skip connections which skips one or more layers in the network, a technique that was also used by (Srivastava et al., 2014). The residual block was refined (He et al., 2016) and a pre-activation residual block with identity transformation was employed (He et al., 2016) wherein the gradients jump to shortcut connections to reach to any other previous layer. By reason of this, ResNet becomes more popular in research community and they came up with many variants of it, namely, ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164 and ResNet-1202, ResNeXt. In our study Resnet50 is used because of its ability to work well on RS data. ResNet50 is 50 layers deep in nature consisting of 5 phases of convolutional layers. Firstly, 7 × 7 convolution with stride two is employed which is succeeded by the pooling layers and then three identity blocks. The last layer is the average pooling layer generating thousand feature maps.

2.2.4. DenseNet201

Dense Convolutional network, DenseNet, was jointly designed by Tsinghua university, Cornwell university and Facebook AI research (FAIR) (Huang et al., 2017). As compared to ResNet and pre-activation ResNet, DenseNet achieved higher accuracy with a smaller number of parameters as a result of dense connectivity. In DenseNet, firstly the input image is convolved with 16 output channels and given to the dense block. In each dense block, all layers are directly connected to every other layer in the block in a feed forward manner. Each layer gathers knowledge from all precursory layers and gives its own output to all following layers. At every layer all the feature maps collected from previous layers are considered separately, concatenated in a single tensor

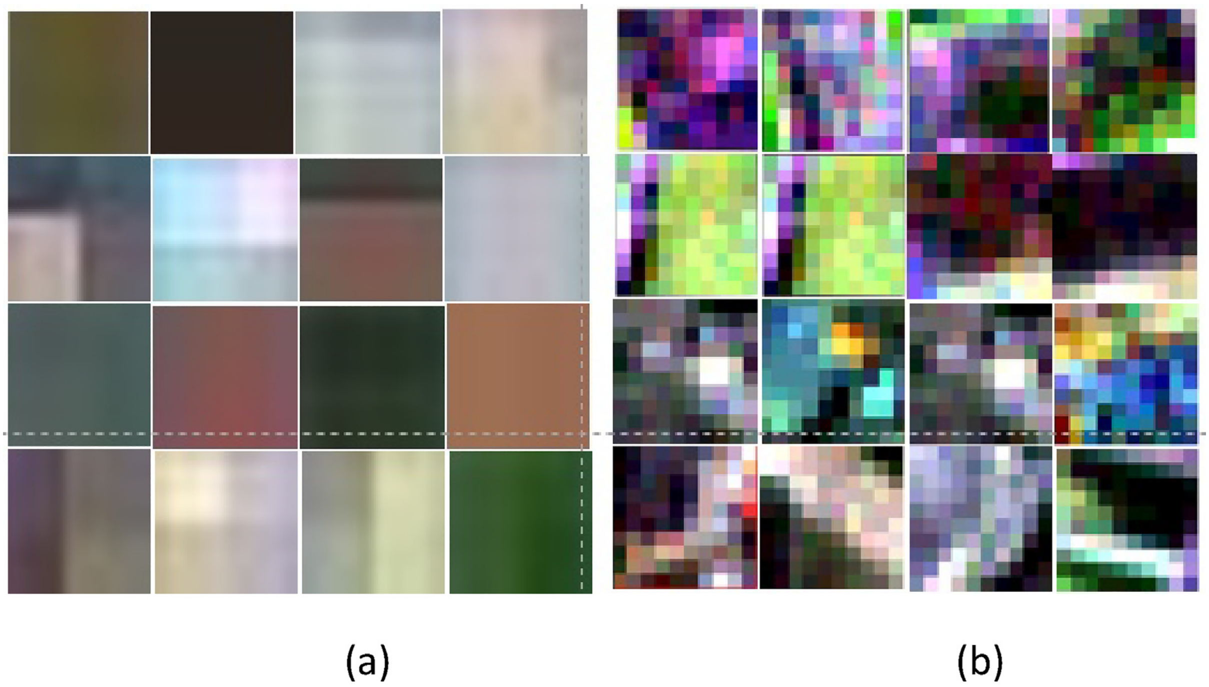


Fig. 4. Example zoomed images of CaneSat dataset in (a) jpeg (b) tif formats.

(Huang et al., 2017) given as an input to the composite function. The composite function constitutes three operations, batch normalization (BN), ReLU and 3 × 3 convolutions. All densely connected dense blocks are joined through transition layers which performs convolution and pooling. Transition layer is made up of a batch normalization layer followed by a 1 × 1 convolutional layer succeeded by a 2 × 2 average pooling layer. Another layer introduced prior to every 3 × 3 convolutional layer in DenseNet is a bottleneck layer, codenamed DenseNet-B, which is designed as BN, ReLU and 1 × 1 convolution layer followed by BN, ReLU and 3 × 3 convolution layer. DenseNet sets the hyper-parameter, referred to as growth rate that defines the amount of new input at each layer, relatively small as each layer in the network has entire network's information. Final dense block is connected to a global average pooling layer and finally a softmax classifier is connected. DenseNet201 is evaluated in this work due to their good performance on remotely sensed data for classification task.

All the above mentioned and other existing deep CNNs can be employed in three different strategies, namely, full training, fine tuning and CNN as FE which are briefly discussed below:

2.2.5. Full training

Full training refers to train a ConvNet right from scratch by assigning random values to kernel weights and is one of the best modalities to get network tuned precisely for a specific dataset. Though it creates accurate features specific to the target dataset and full control on network parameters, it necessitates a large dataset to converge the network which demands huge computational cost, recourses (Bengio et al., 2009) and faces overfitting problem. In full training, we can either utilize the existing network configuration with randomly initialization of its weights or can model a new network architecture by setting all its components, namely, number of processing units, convolutional layers, pooling layers, FC layers, activation functions, number of epochs, learning rate, weight decay, type of normalization and regularization techniques.

2.2.6. Fine tuning

Fine tuning is a concept, where information acquired by a network model through training process on a task is utilized to train the model to perform another analogous task. It relies on a property of learning low-level features that resemble color blob or edge detectors, Gabor filters and mainly it is independent on the training dataset. Later layers extract features which are specific to the problem. This makes it more suitable option when the training dataset is large enough but insufficient to train a network from scratch (Nogueira et al., 2017). We employed fine tuning on the CaneSat dataset in this study.

2.2.7. Feature extractors

Encoding perspicacious features from visual data is one of the major phases in computer vision tasks, RS being no exception for this. Due to particularities in RS data, many of the long-established methods like color histograms, correlograms (Kumar and Bhatia, 2014), BoVW (Tsai, 2012) are not simply applied in RS domain and hence, it is still an open research problem (Nogueira et al., 2015). CNNs can be used as arbitrary FE in DL framework wherein training images are propagated through the layers of the pre-trained network and taken output after any pre-defined layer as deep features. These features can then be provided to any other linear classifiers to perform classification task. Deep features trained on ImageNet produced astounding results in many visual recognition tasks confirms the extracted deep features will work well for dataset of interest other than the ImageNet dataset. Accordingly, we utilized the above discussed four pre-trained models as FE in our present study.

As observed from literature, deep CNNs reveal superiority in performance over other traditional machine learning approaches in land use, land cover and crop classification. Therefore, we use DL models for the classification of sugarcane.

2.3. Methodology

The focus of this work is to develop a CNN model to classify sugarcane crops from other land covers on the created CaneSat dataset and to investigate the power of the four selected pre-trained CNNs on the target CaneSat dataset. The selected pre-trained convolutional networks' performances were evaluated by employing two modalities i.e., fine tuning the layer before classification and CNNs as FE. All the experiments for 2D CNN model were carried out on an Intel(R) Xeon(R) CPU E3-1271 v3 @3.60 GHz with 32 GB RAM. A 64-bit 7th generation Intel core i7 processor with 2.80 GHz operating speed was used for the deep networks' training and experimentational evaluation. The quantitative evaluation of various models used in this study is performed using overall accuracy as metric that is derived from the confusion matrix. The overall accuracy is calculated as the percentage of accurately classified classes of the test dataset. Class level accuracy is evaluated by F1 score which gives harmonic mean of precision and recall. Precision is defined as correctly predicted classes divided by the total number of classes predicted by the model. The recall is the proportion of correctly predicted classes to actual classes.

2.3.1. Architecture of the proposed 2D CNN model

Six bands from Sentinel-2 imagery forms feature vectors are given as the input to the 2D convolutional neural network trained on the CaneSat dataset. The architecture of this 2D CNN model is shown in Fig. 5.

It has six layers comprising of three convolutional layers, one max pooling layer, one fully connected layer and a softmax layer. The kernel size for all three convolutional layers is 3 × 3. The number of filters is started with three for first convolutional layer and increased to six and nine for second and third layer respectively. The size of output feature maps of every convolutional layer is retained, to maintain the original information of the input image patch (padding = 'same'). Pooling layer with 2 × 2 filter size is applied after third convolutional layer, as image patches are small and pooling will not help in the beginning layers. The convolutional layers and pooling layer consider all the image pixels (stride is set to one) into convolution and down sampling operation, respectively. Convolutional layer uses ReLU activation function which is one of the most efficient and powerful activation functions to add non-linearity in the network. It has many advantages such as fast gradient propagation, good control on vanishing gradient and computationally efficient because of not activating all neurons simultaneously and has convergence better than sigmoid. binary crossentropy was used as the loss function. Regularization was adopted through dropout that is calibrated to 0.25 probability value and is attached after the FC layer to avoid overfitting. Finally, the fully connected layer is followed by a softmax layer to form probability distribution from the network's output over two classes.

2.3.2. Transfer learning

Four deep CNNs AlexNet, GoogLeNet, ResNet50 and DenseNet201, originally trained on ImageNet dataset, are chosen based on their popularity in research class and their astonishing performance in RS applications. In this research study, performance of pre-trained deep CNNs on the created CaneSat dataset is evaluated by employing two techniques of transfer learning. In first technique, so-called fine tuning, trained weights of these pre-trained deep CNNs are transferred and only final fully connected layer is fine-tuned with respect to the target dataset as shown in Fig. 6, highlighted in red.

In second technique, pre-trained CNNs are exploited as FE. From the literature survey, it is observed that deep features work well in classification problem compared to the conventional features. We can extract deep features from any of the layers of the deep networks. In the present study, features are retrieved from FC7, FC1000, global average pool and avgpool layers of AlexNet, GoogLeNet, ResNet50 and DenseNet20 respectively and passed on to the SVM classifier by passing the Softmax layer to classify sugarcane and non-sugarcane (Fig. 6, highlighted in green). Literature survey reveals that the SVM is best for crop classification from satellite images. Hence, it is employed as classifier in the transfer learning implementation. The radial basis function is used for SVM kernel with 'C' parameter having value one and gamma is set to value 0.6. The batch size is fixed to 128. At the 14th epoch model has converged with learning rate of 0.001. The optimization was achieved by advanced adaptive moment (Adam) optimizer and the loss function was selected as cross-entropy.

2.3.3. RF and SVM ML techniques

The research utilized two widely employed machine-leaning approaches for crop classification, RF and SVM, to compare them with DL techniques. In fact, RF and SVM methods have been commonly used as baseline models for the DL approach to RS classification tasks (Mahdianpari et al., 2018; Makantasis et al., 2015). The RF is an ensemble of multiple decision trees working on bootstrap data (Breiman, 2001). The SVM is a statistical learning classifier tries to find optimal hyperplane between classes with the help of kernel functions (Cortes and Vapnik, 1995). The classification was conducted using Sentinel-2 time series data and was implemented in R language.

3. Results and discussion

Here, we exhibit and discuss experimentation results of the research work. Firstly, the success of the 2D CNN model for classifying sugarcane is explored. Next, the specified pre-trained networks are capable of generalizing RS data for crop classification is tested, employing two techniques using the CaneSat dataset. One of the aspects of this evaluation is to find how well transfer learning performs on a small RS dataset as compared to the available big RS datasets such as UCM, BCD and RS19.

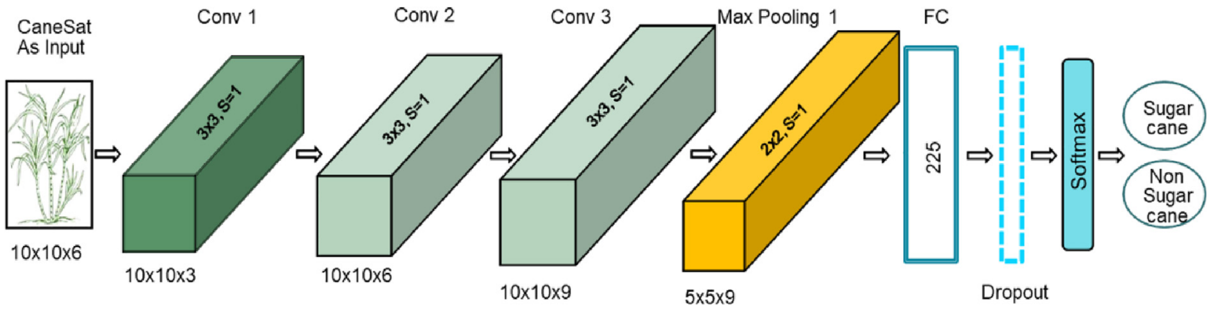
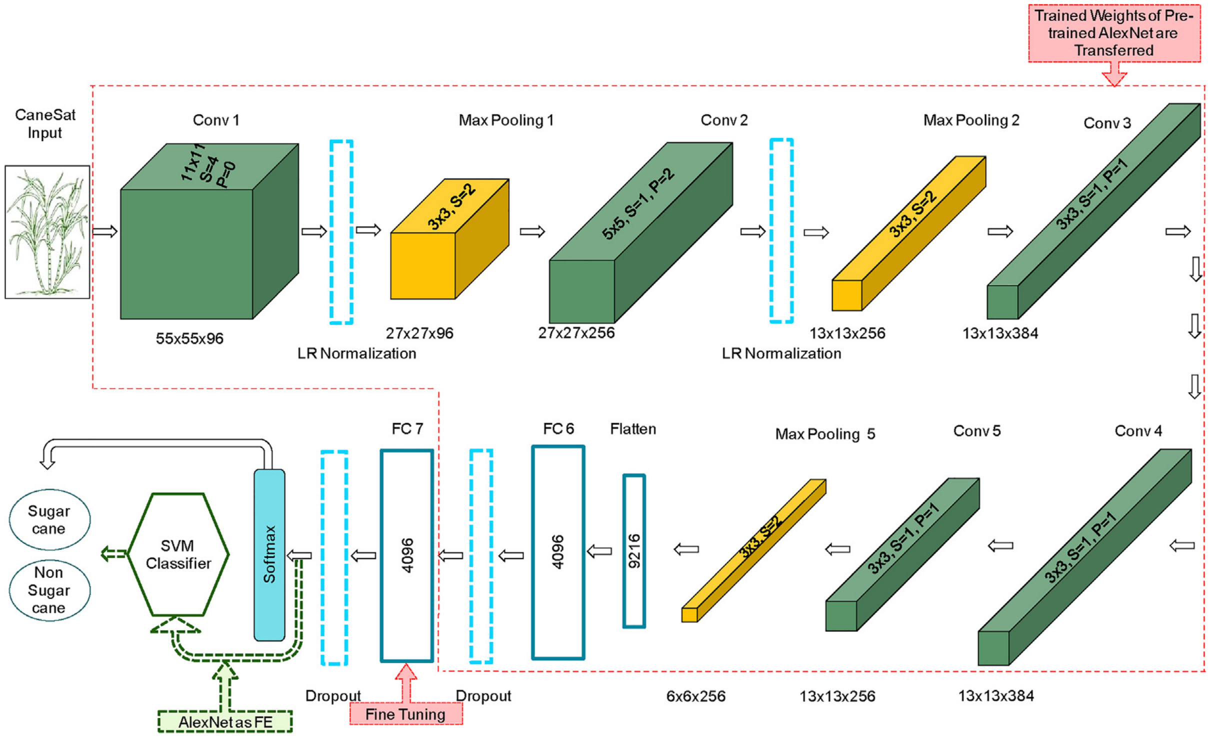


Fig. 5. 2D convolutional neural network on the CaneSat dataset.



(a) AlexNet

Fig. 6. CNNs with fine tuning (highlighted in red) and as FE with SVM classifier (highlighted in green).

3.1. 2D CNN model's evaluation for sugarcane classification

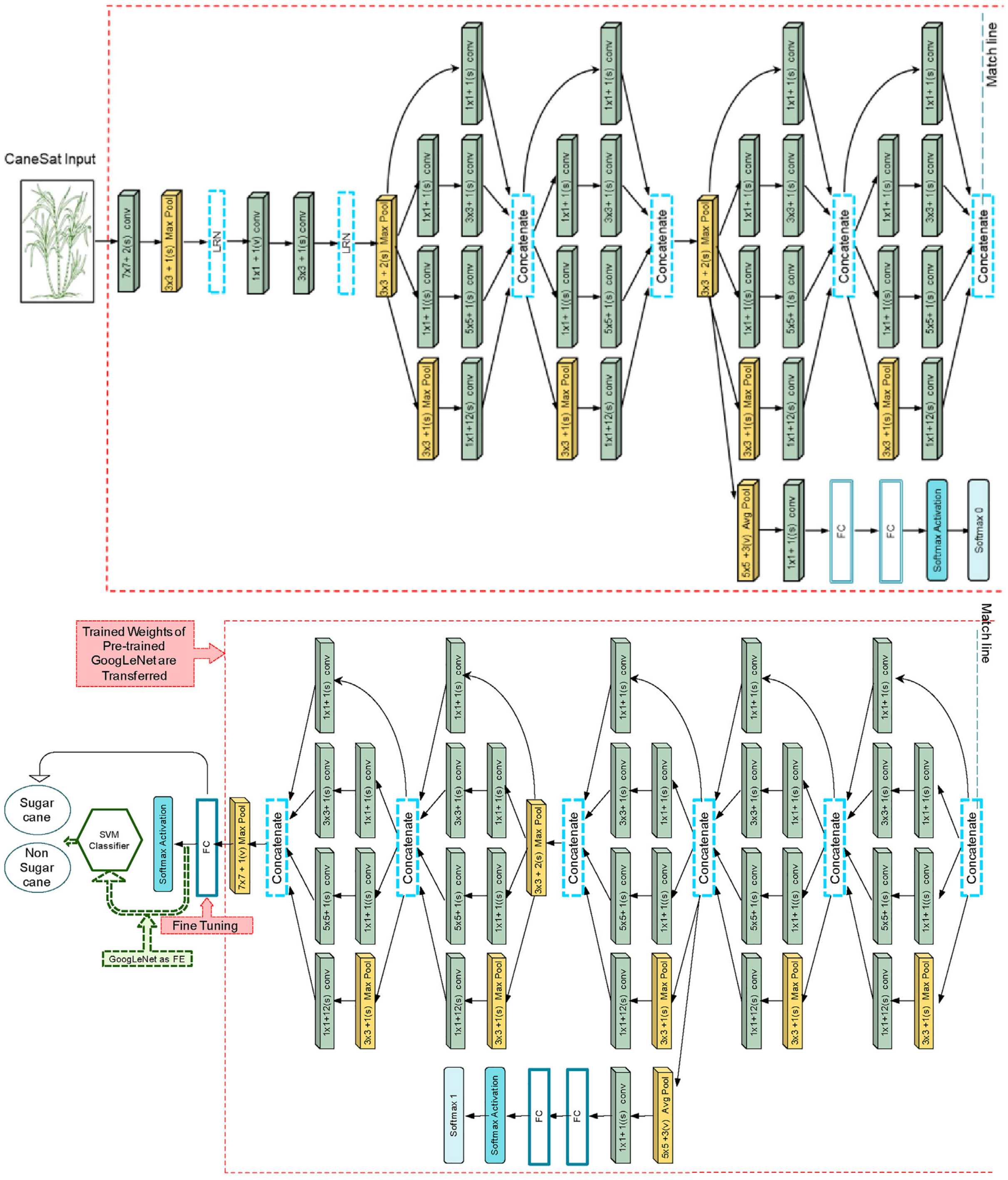
Several experiments have been carried out to investigate the performance of the model on the newly created 'CaneSat' dataset. The CaneSat dataset was split into 70% for training and 30% for testing which observed good accuracy. Further, the training dataset is divided into 70% for training and 30% for validation. Batch size is calibrated to 128 which contributes in accuracy boost up. The trainable parameters of the model are 3110. The training and the testing datasets are normalized after undergoing through atmospheric corrections. Another transformation technique in neural network, called standardization plays an important role in the classification results. In our case the training and the testing datasets have undergone the standardization phase through StandardScaler function. The learning rate was set to 0.001. The model has converged in 14 epochs with the Adam optimizer. Adam was opted because of its faster convergence rate than stochastic gradient descent (SGD) and RMSProp (Kingma and Ba, 2014). In our scenario, various experiments were carried out to investigate Adam, SGD and RMSProp optimizers. SGD performs closest to Adam with the difference of 3–4% less accuracy compared to Adam. With this configuration, the 2D CNN model trained on the CaneSat dataset reveals significant performance for sugarcane crop classification with 88.46% overall accuracy. The validation accuracy reached up to 84.65% concludes that the model possesses good generalization ability. Fig. 7 depicts model's training and validation accuracy and training and validation loss.

The sugarcane crop was correctly identified with an accuracy of 86.99% and non-sugarcane areas with high accuracy of 90.12%. The time taken to train the model is 1.80 min (Table 2). In literature, LULC classification from RS data are more exploited using DL framework than crop type mapping and achieved promising results with accuracy above 90% (Penatti et al., 2015). 1D CNN was successfully exploited by Helber et al. (2019) for crop classification by attaining 82.41% accuracy. In LULC classification few crops are considered in their studies, but we found fewer studies considering sugarcane crop (Ienco et al., 2019). We accomplished relatively good accuracy with the 2D CNN model on CaneSat dataset containing fewer numbers of samples compared to other large-scale datasets that confirms the potential of CNN in sugarcane classification.

3.2. Pre-trained models' evaluation for sugarcane classification

Confusion matrix defines confusion among classes is used to evaluate all four deep CNNs. Confusion matrices of deep CNNs fine tuning and as FE with SVM classifier are shown in Fig. 8. As illustrated by the confusion matrices (Fig. 8), all CNNs achieved notable high accuracy in classifying the sugarcane class. Among all other networks, GoogLeNet found less confusion (when fine-tuned 8.05% and 1.53% as FE) in classifying the sugarcane class. However, it has more confusion of 44.05% when fine-tuned and 55.06% as FE in classifying the non-sugarcane class. An uncertainty of 33.04%, 31.71%, 20.26% in fine-tuning method and 22.90%, 23.79%, 16.74% in FE method exists in classifying the non-sugarcane class for AlexNet, ResNet50 and DenseNet201 respectively. Non-sugarcane class was comparatively correctly classified by DenseNet201 in the both modalities with accuracy of 83.25% as fine-tuned and 79.73% accuracy as FE. Both the classes' F1 scores are compared in Fig. 9 for the transfer learning techniques, fine tuning and CNNs as FE.

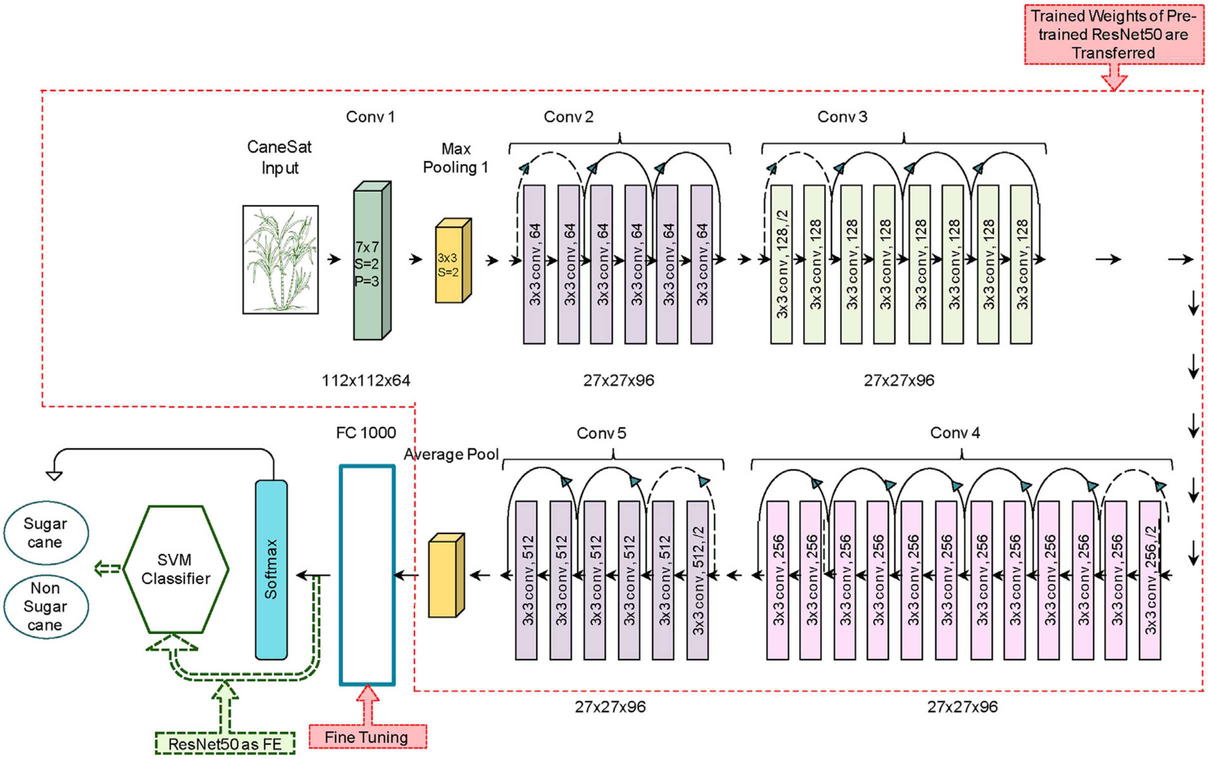
2D CNN was executed on a CPU whereas transfer learning was implemented on a GPU. Table 2 presents comparison of model's performance regarding the required training time and achieved accuracy by the pre-trained models on the CaneSat dataset. From



(b) GoogLeNet

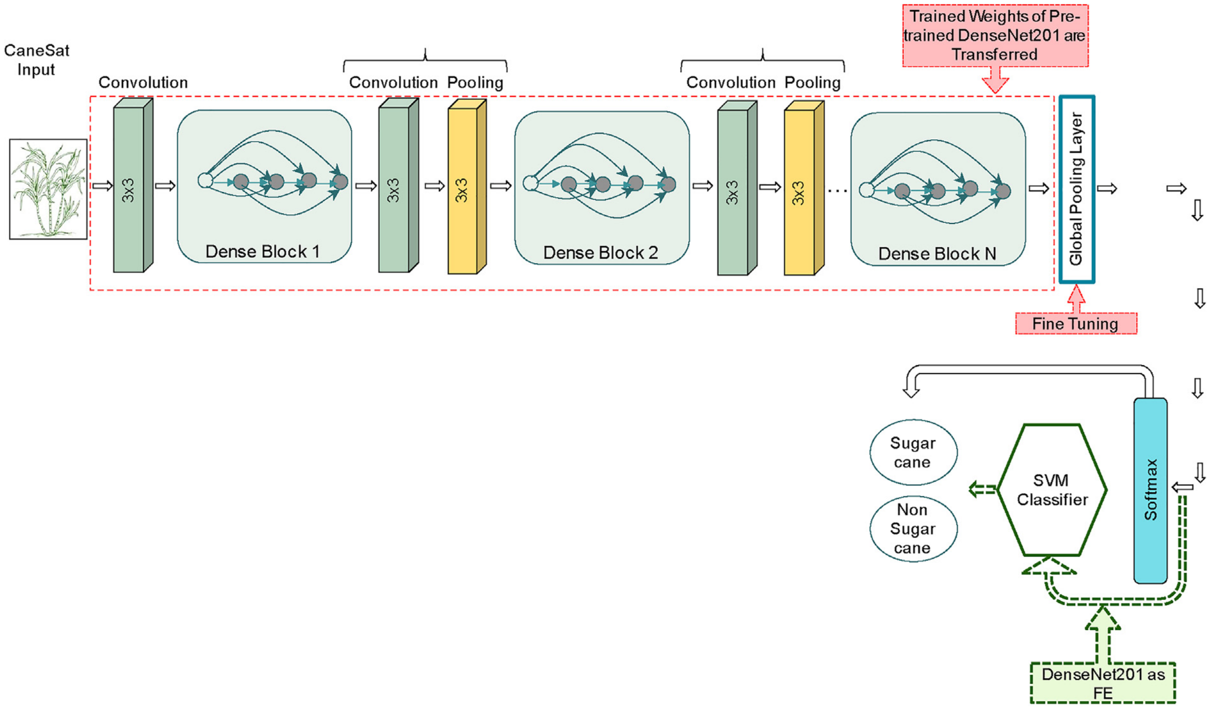
Fig. 6 (*continued*)

the obtained results of the study, it is observed that AlexNet converged in less training time than all other three networks. It is because of its smaller number of layers and parameters and giving better accuracy than GoogLeNet. As observed from Table 2 ResNet50 and DenseNet201 required comparatively more time than AlexNet and GoogLeNet due to the deep nature of their structure. One of the most attentive features of the achieved results in this present study is that deep CNNs used as FE produced better results than fine tuning modality for all CNNs except GoogLeNet. In fact, a significant difference exists between our dataset and the original dataset which these CNNs were trained on. An optimal approach will also be a fine-tuning approach if low level features of



(c) ResNet50

Fig. 6 (*continued*)



(d) DenseNet201

Fig. 6 (*continued*)

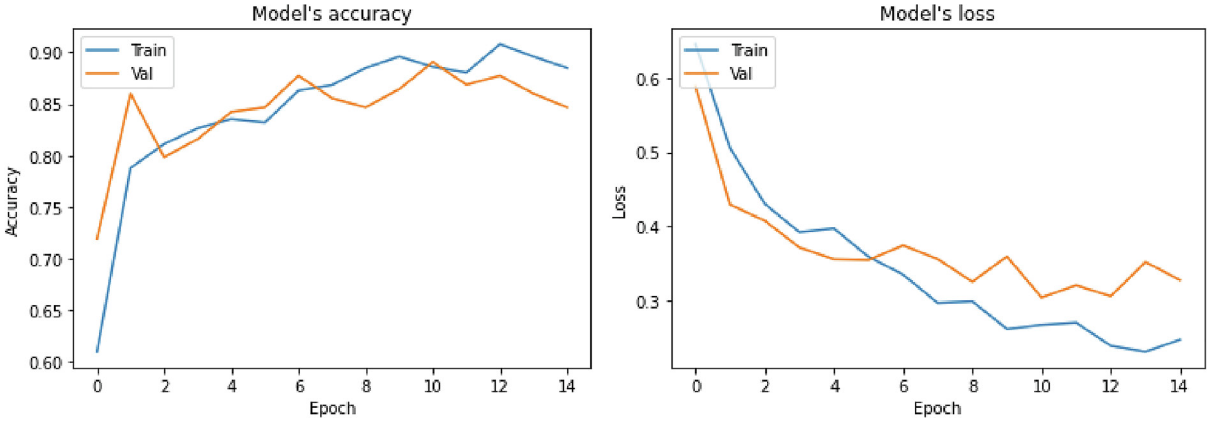
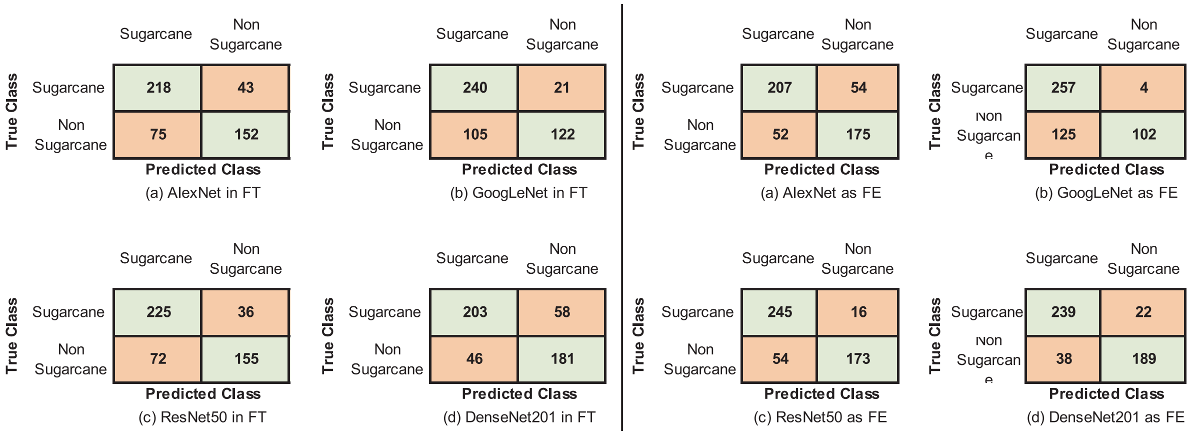


Fig. 7. 2D CNN's (a) training and validation accuracy (b) training and validation loss.

Table 2
Comparison of models' performance.

Model	Full Training		Fine Tuning		Deep CNNs as FE	
	Training-time (minutes)	Accuracy (%)	Training-time (minutes)	Accuracy (%)	Training-time (minutes)	Accuracy (%)
SVM	0.20	80.38				
RF	0.21	84.00				
2D CNN	1.80	**88.46**	–	–	–	–
AlexNet	–		7.28	75.81	7.38	78.28
GoogLeNet	–		30.40	74.18	14.41	73.56
ResNet50	–		34.60	77.86	36.24	**85.65**
DenseNet201	–		181.11	78.68	177.17	**87.70**



True Class	Sugarcane	Non Sugarcane
Sugarcane	218	43
Non Sugarcane	75	152

Predicted Class
(a) AlexNet in FT

True Class	Sugarcane	Non Sugarcane
Sugarcane	240	21
Non Sugarcane	105	122

Predicted Class
(b) GoogLeNet in FT

True Class	Sugarcane	Non Sugarcane
Sugarcane	207	54
Non Sugarcane	52	175

Predicted Class
(a) AlexNet as FE

True Class	Sugarcane	Non Sugarcane
Sugarcane	257	4
Non Sugarcan e	125	102

Predicted Class
(b) GoogLeNet as FE

True Class	Sugarcane	Non Sugarcane
Sugarcane	225	36
Non Sugarcane	72	155

Predicted Class
(c) ResNet50 in FT

True Class	Sugarcane	Non Sugarcane
Sugarcane	203	58
Non Sugarcane	46	181

Predicted Class
(d) DenseNet201 in FT

True Class	Sugarcane	Non Sugarcane
Sugarcane	245	16
Non Sugarcane	54	173

Predicted Class
(c) ResNet50 as FE

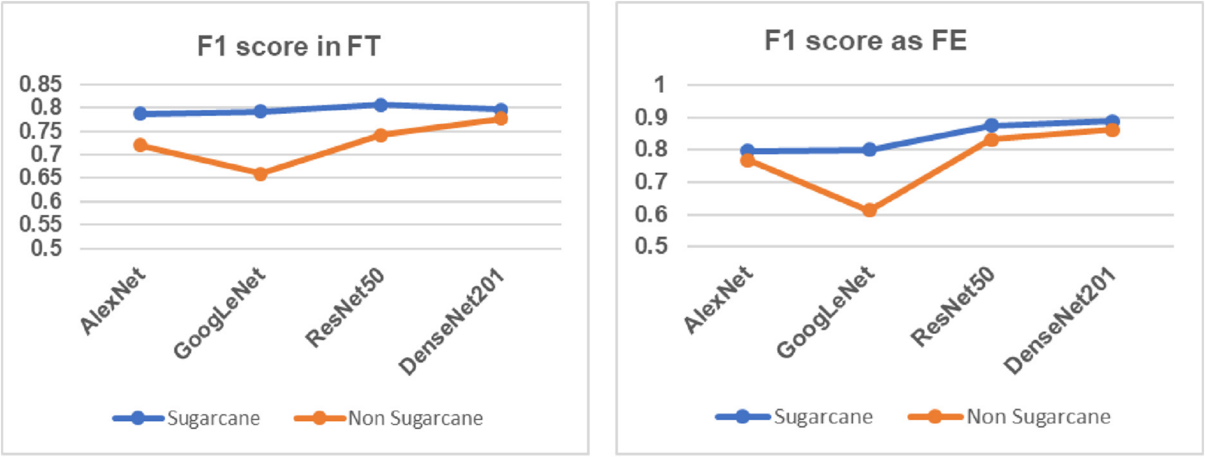
True Class	Sugarcane	Non Sugarcane
Sugarcane	239	22
Non Sugarcan e	38	189

Predicted Class
(d) DenseNet201 as FE

Fig. 8. Confusion matrices of four deep networks in fine tuning (FT) and as FE with SVM classifier (highlighted in green).

the interest dataset were found to be close to those on which the CNNs were trained on.

However, the low-level features such as color, texture and contours of ImageNet data set objects vary to some degree from the objects of our dataset. It may explain why the fine-tuning method acquires less accuracy than used as FE. In our case, a layer before classification layer is fine-tuned, however, the results may differ in this method by changing the number of layers to be fine-tuned. Fine-tunned AlexNet performs as closely as other advanced networks. Features extracted from first or second layer of the AlexNet for high resolution RS dataset (UCM) achieved remarkable accuracy (95%) (Hu et al., 2015). GoogLeNet's performance remains good in its fine-tuning strategy with 74.18% accuracy over its use as feature extractor with SVM classifier, resulting in 73.56% accuracy. In our study, deep CNNs are used as FE and features are extracted from a layer before the classification (softmax) layer from the original network. So, the classification process of the original network is changed. The emphasis is on discovering how well deep features function on datasets that are different in nature from the original datasets



(a) With fine-tuned modality

(b) Deep CNNs as FE with SVM classifier

Fig. 9. Comparison of F1 score of AlexNet, GoogLeNet, ResNet50 and DenseNet201.

used to trained deep networks. Here, the deep features extracted using CNNs are fed to the SVM classifier rather than a softmax layer. The softmax layer is essentially a logistic regression generalization; it operates on the concept of assigning probabilities to each class label. Crop classification requires a comprehensive methodology to distinguish sugarcane with other crops and land covers that cannot be done effectively by a regression methodology. Therefore, we used features extracted using CNNs and fed them as input to an SVM classifier. SVM classifier is essentially designed to resolve classification problems and is the most proven classifier with fewer samples, without losing precision in RS image classification tasks including LULC and crop classification. The result on CaneSat dataset demonstrates the capacity of CNNs extracted features to generalize to the multispectral RS data for crop classification. GoogLeNet performance in this study found to be lower than AlexNet, is close to the study results reported in Nogueira et al. (2017) when both CNNs are used as FE with RBF SVM for UCM, RS19 and BCS datasets (AlexNet > 93% accuracy and GoogLeNet with 92.80% for UCM dataset). Features extracted by GoogLeNet and classified by softmax layer instead of SVM reached 90.75% accuracy and when fine-tuned 84.00% accuracy on BCS dataset (97.10% and 94.38% for UCM dataset) found by Castelluccio et al. (2015). ResNet50 and DenseNet201 trained on CaneSat dataset achieved significant high overall accuracy of 85.65% and 87.70%, respectively, when used as FE with SVM classifier compared to fine tuning method. Using fine tuning method ResNet50 and DenseNet201 gained accuracies 77.86% and 78.68%, respectively. These findings contrast with the findings recorded by Mahdianpari et al. (2018) for wetland classification using green, red and infrared bands for ResNet50 (93.00%) and DenseNet121 (86.90%). DenseNet121 under performed in their analysis as against VGG16, VGG19, InceptionV3, ResNet50 and InceptionResNetV2.

3.3. RF and SVM classification

The two ML techniques RF and SVM have been implemented on the CaneSat dataset to compare deep networks' ability in crop classification. The input given to these classifiers are Sentinel-2 time series during Oct. 2018 – May 2019. As shown in Table 2, RF and SVM classified sugarcane and non-sugarcane with relatively low overall accuracies (84.00%, 80.38%) than our 2D CNN.

4. Conclusion

This research work addresses the problem of sugarcane crop classification. For this endeavor, a novel dataset is prepared from the open and freely available Sentinel-2 A/B satellite images dispensed by ESA's Copernicus program. The presented CaneSat dataset consists of sugarcane monthly samples covering all the stages from one month to 12 months along with other land cover samples. The dataset comprises overall 1627 samples labeled in two classes, sugarcane and non-sugarcane. The key emphasis here is to cover all the growing phases of sugarcane crop to train the network. Dataset images are labeled and available in two formats i.e. jpg and tif format. It covers four talukas of the Karnataka state of the India. All the geo-tiff images are prepared by using six spectral bands of the remotely sensed images and are geo-referenced. A model using 2D CNN is developed for this dataset to classify sugarcane and the other land covers. Further, deep CNNs, namely, AlexNet, GoogLeNet, ResNet50 and DenseNet201 performances are evaluated when fine-tuned and used as FE with SVM classifier on this dataset. The results illustrate that the 2D CNN model achieves 88.46% accuracy with six layers in the network. On the other hand, all deep networks perform well when used as FE. We achieved 81.76% and 74.66% average f1 scores for the sugarcane and the non-sugarcane class respectively and an average overall accuracy of 78.97%, despite the challenges of small sugarcane sample size and a smaller number of images in the dataset as compared to the popular RS LULC datasets such as UCMerced and EuroSat. The GoogLeNet (74.18% and 73.56%) is little incompetent than all other models. The DenseNet201 (87.70%) followed by the ResNet50 (85.65%) demonstrate noticeable accuracy. Fine tuning of AlexNet, ResNet50, DenseNet201 and GoogLeNet before softmax layer gained 75.81%, 77.86%, 78.68% and 74.18% accuracies, respectively.

In summary, ResNet50 extracts global features using skip connection so has powerful representation and DenseNet201 extracts features from all complexity levels which is possible due to the dense construction in the network. Therefore, ResNet50 and DenseNet201 gained significant accuracy in classification task, which is confirmed in our case as well, when used them as FE along with the SVM classifier.

The presented CaneSat dataset is a great step to leverage the openly available satellite images in the agriculture domain specifically in monitoring vegetations at a regional level. Though the

CaneSat dataset is different than the Imagenet dataset on which the deep CNNs were trained on, the results of the present research work exhibit the generalization ability of these pre-trained CNNs in classifying the multispectral remotely sensed data.

It is intended to work in future on the quality of the sugarcane such as healthy, biotic and abiotic stressed crops. Further, to investigate the potential of deep networks on SAR images for sugarcane and non-sugarcane classification.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Shri Sameer Somaiya for his support and the staff of GBL and KIAAR, Sameerwadi, Karnataka, India for their efforts in collecting ground truth data of crop plots.

References

Atkinson, P.M., Tatnall, A.R.L., 1997. Introduction neural networks in remote sensing. Int. J. Remote Sens. 18, 699–709.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. Deepsat: a learning framework for satellite imagery, in. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, p. 37.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523.

Bengio, Y. et al., 2009. Learning deep architectures for AI. Found. trends®in Mach. Learn. 2, 1–127.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks. arXiv Prepr. arXiv1508.00092.

Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. Proc. IEEE 105, 1865–1883.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: In Proceedings of the 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255.

He, K., Sun, J., 2015. Convolutional neural networks at constrained time cost, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5353–5360.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Helber, P., Bischke, B., Dengel, A., Borth, D., 2019. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 2217–2226.

Hu, F., Xia, G.-S., Hu, J., Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 7, 14680–14707.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Ienco, D., Gaetano, R., Dupaquier, C., Maurel, P., 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geosci. Remote Sens. Lett. 14, 1685–1689.

Ienco, D., Interdonato, R., Gaetano, R., Minh, D.H.T., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding, in. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678.

Interdonato, R., Ienco, D., Gaetano, R., Ose, K., 2019. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. ISPRS J. Photogramm. Remote Sens. 149, 91–104.

Kingma, Diederik P and Ba, Jimmy, 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kumar, G., Bhatia, P.K., 2014. A detailed review of feature extraction in image processing systems. In: in: 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 5–12.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: In Proceedings of the Advances in neural information processing systems, pp. 1097–1105.

Krizhevsky, A., 2009. Learning Multiple Layers of Features from Tiny Images. Ph.D. dissertation.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al., 1998. Gradient-based learning applied to document recognition. Proc. IEEE. 86, 2278–2324.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sens. 10, 1119.

Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N., 2015. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4959–4962.

Nogueira, K., Miranda, W.O., Dos Santos, J.A., 2015. Improving spatial feature representation from aerial scenes by using convolutional networks. In: in: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images, pp. 289–296.

Nogueira, K., Penatti, O.A.B., dos Santos, J.A., 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit. 61, 539–556.

Penatti, O.A.B., Nogueira, K., Dos Santos, J.A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 44–51.

Ponti, M., Chaves, A.A., Jorge, F.R., Costa, G.B.P., Colturato, A., Branco, K.R., 2016. Precision agriculture: Using low-cost systems to acquire low-altitude images. IEEE Comput. Graph. Appl. 36, 14–20.

Rezaee, M., Mahdianpari, M., Zhang, Y., Salehi, B., 2018. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 3030–3039.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv Prepr. arXiv1409.1556.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Tsai, C.-F., 2012. Bag-of-words representation in image annotation: A review. ISRN Artif, Intell, p. 2012.

Virnodkar, S.S., Pachghare, V.K., Patil, V.C., Jha, S.K. 2019a Application of Machine Learning on Remote Sensing Data for Sugarcane Crop Classification: A Review, In: Fong S., Dey N., Joshi A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 93. Springer, Singapore.

Virnodkar, S.S., Pachghare, V.K., Patil, V.C., Jha, S.K.S., 2019. Performance Evaluation of RF and SVM for Sugarcane Classification using Sentinel 2 Time Series NDVI, in. Springer AISC (In press).

Virnodkar, S.S., Pachghare, V.K., Patil, V.C., Jha, S.K., 2020. Remote sensing and machine learning for crop water stress determination in various crops: a critical review. Precis. Agric., 1–35

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A large-scale dataset for object detection in aerial images, in. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3974–3983.

Xia, G.-S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X., 2017. AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Trans. Geosci. Remote Sens. 55, 3965–3981.

Yan, J., Wang, H., Yan, M., Diao, W., Sun, X., Li, H., 2019. IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery. Remote Sens. 11, 286.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification, in. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 270–279.

Yang, C., Everitt, J.H., Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. Comput. Electron. Agric. 75, 347–354.

Zheng, H., Zhou, X., He, J., Yao, X., Cheng, T., Zhu, Y., Cao, W., Tian, Y., 2020. Early season detection of rice plants using RGB, NIR-GB and multispectral images from unmanned aerial vehicle (UAV). Comput. Electron. Agric. 169, 105223.

Zhou, W., Newsam, S., Li, C., Shao, Z., 2018. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. ISPRS J. Photogramm. Remote Sens. 145, 197–209.