



DenseResUNet: An Architecture to Assess Water-Stressed Sugarcane Crops from Sentinel-2 Satellite Imagery

Shyamal S. Virnodkar^{1,2*}, Vinod K. Pachghare¹, Virupakshagouda C. Patil³, Sunil Kumar Jha³

¹ Department of Computer Engineering & IT, College of Engineering, Savitribai Phule Pune University, Pune 411005, India

² K J Somaiya Institute of Engineering and Information Technology, Mumbai University, Sion 400022, Mumbai, India

³ K J Somaiya Institute of Applied Agricultural Research (KIAAR), Bagalkot District, Sameerwadi 587316, Mudhol Taluk, India

Corresponding Author Email: shyamal@somaiya.edu

<https://doi.org/10.18280/ts.380424>

ABSTRACT

Received: 22 March 2021

Accepted: 25 July 2021

Keywords:

sugarcane crop, Sentinel-2, deep learning, crop water stress, DenseResUNet

A single most immense abiotic stress globally affecting the productivity of all the crops is water stress. Hence, timely and accurate detection of the water-stressed crops is a necessary task for high productivity. Agricultural crop production can be managed and enhanced by spatial and temporal evaluation of water-stressed crops through remotely sensed data. However, detecting water-stressed crops from remote sensing images is a challenging task as various factors impacting spectral bands, vegetation indices (VIs) at the canopy and landscape scales, as well as the fact that the water stress detection threshold is crop-specific, there has yet to be substantial agreement on their usage as a pre-visual signal of water stress. This research takes the benefits of freely available remote sensing data and convolutional neural networks to perform semantic segmentation of water-stressed sugarcane crops. Here an architecture 'DenseResUNet' is proposed for water-stressed sugarcane crops using segmentation based on encoder-decoder approach. The novelty of the proposed approach lies in the replacement of classical convolution operation in the UNet with the dense block. The layers of a dense block are residual modules with a dense connection. The proposed model achieved 61.91% mIoU, and 80.53% accuracy on segmenting the water-stressed sugarcane fields. This study compares the proposed architecture with the UNet, ResUNet, and DenseUNet models achieving mIoU of 32.20%, 58.34%, and 53.15%, respectively. The results of this study reveal that the model has the potential to identify water-stressed crops from remotely sensed data through deep learning techniques.

1. INTRODUCTION

Swift increase in the world's population pressurizes the availability of fresh water. The population largely survives on agricultural products. Agriculture stood first (near 70%) in the freshwater consumption sectors among domestic, industries, and civil sectors. However, the increase in global warming day by day hinders freshwater availability. With this limited freshwater, sustainable agriculture and food security become significant issues for the next generation. Water deficiency is not limited to any growing stage of any crop, but it may arise continuously over the entire crop lifecycle. It may be more or less in a particular stage; however, it inhibits the overall growth and is the primary cause of the reduction in the yield (more than 50%) of the crop, which causes significant loss to the farmers, industries, and country's economy. The crop yield responds to water deficiency differently in different crops and is also dependent on surrounding climatic conditions. The determination of water stress in crops is a great challenge to schedule adequate irrigation to ensure food security. In addition, the agriculture domain faces the problem of producing more food from this limited water supply. In this study, an attempt is made to assess the water stress of the sugarcane crop in Karnataka, India.

Sugarcane (*Saccharum officinarum*) is a widely grown and

one of the most important crops cultivated in India under varying conditions of the atmosphere, soil types, rainfall, temperature regimes, and mainly water availability. Sugarcane crop cultivation requires a massive amount of water ranges from 1200 to 3500 mm (1200-1800 mm in subtropical and 1600-2700 mm in tropical zones) that depends on the season (time of plantation), crop type, crop duration, yield, climate, and soil properties. The average rainfall in Bagalkot and Belgavi (study area) is 545 mm and 808 mm, respectively. Hence, efficient water management is much needed to better yield in sugarcane cultivation which is possible only through timely and accurate assessment of the crop water stress.

This paper proposes a novel architecture to investigate the potential of combining encoder-decoder structure, ResNet, and DenseNet architectures to gain the advantages of these techniques for identifying high water stress, medium water stress, and low water stress sugarcane from the satellite imagery. The proposed architecture followed UNet's overall encoder decoder structure with significant changes in simple convolution and decoding paths. Residual units connected through dense connection are utilized in place of the plain convolutional layers. This study also investigated an attention module that concentrates on the dataset's significant features. This paper aims to gain acceptable segmentation results for water-stressed sugarcane plots. The proposed architecture is

evaluated on the developed dataset for water-stressed sugarcane crops.

The remaining paper is organized as follows. Section 2 gives a literature survey. Section 3 briefs the methodology used in this research study. Section 4 presents the experiments conducted with the dataset used. The investigation results and comparison of the model are given in section 5, followed by the research conclusion in section 6.

2. LITERATURE SURVEY

Various methods to detect crop water stress have been reported in the literature over the years vary from in situ soil moisture measurement, methods based on plant responses, remote sensing (RS) based methods to machine learning (ML) based methods [1]. Earlier, crop water stress detection was carried out by taking dry weight and fresh weight of the crop leaves, which were very time consuming, destructive, laborious techniques and cover a tiny region of the farming. Then, a practical, non-destructive technology, the RS technology covering a wide area, showed superior performance over the laboratory-based techniques for water stress detection in crops. The spectral indices of multispectral satellite images, especially crop water stress index, hyperspectral satellite images, witnessed great success in detecting water stress in crops.

Recent years bring amelioration in deep learning (DL) methods (deep neural networks), especially convolutional neural networks (CNNs). CNN gained popularity because of achieving state-of-the-art results in many computer vision tasks such as image classification [2-4], scene recognition [5], image annotation and captioning [6, 7], handwritten digits classification, object detection [8], and semantic segmentation [9]. A review has been presented by Singh [10] on biotic and abiotic stress detection using CNN models such as AlexNet, ZFNet, GoogLeNet, VGGNet, ExceptionNet, InceptionNet, and ResNet in various fruits and vegetables. Stress and non-stress water condition of maize, soybean, and okra have been identified using AlexNet, GoogLeNet, and Inception V3 [11]. A review on analyzing crop biotic and abiotic stress using deep learning approaches has been reported by Gao et al. [12]. Among other approaches, semantic segmentation has been evaluated for a few crops' water stress determination. It is also called pixel-based segmentation, or per-pixel prediction problem that assigns a class to every pixel of an image. CNNs are promising for it. An extension of CNN, the fully convolutional network (FCN) [9] became one of the most successful techniques for image segmentation. FCN removed a fully connected layer, and the architecture is made up of locally connected layers, convolutional and pooling layers. Removal of this dense layer lessens the parameters and, in turn, the computation time. FCNs introduced up-sampling layers to conventional CNNs to map features to features and obtain spatial resolution at the output end. This up-sampling layer leads to having an arbitrary size of an image at the input end. In addition, identity mapping was introduced, connecting contracting and expansive paths that deal with the fine-grained spatial content loss that occurred by the pooling layer in the contracting path. Over the past few years, many studies have been continuously contributed to the architectures to improve the semantic segmentation using FCNs. FCNs encoders are enhanced [13, 14], and various variants of expansive paths resulted in better segmentation [15, 16]. A breakthrough

architecture, UNet, was proposed by Ronneberger et al. [15] with identity mapping against in FCNs, which concatenates local features from contracting layers to the corresponding expansive layers that proved to be superior against state-of-the-art to segment biomedical images. After the successful performance of UNet on medical image segmentation, many researchers reported noticeable advancement in the original UNet architecture such as H-DenseUNet [17], Dense-InceptionUNet [18], ResUNet++ [19] on the medical image datasets MICCAI 2017 Liver Tumor Segmentation (LiTS) Challenge and 3DIRCADb, lung segmentation in CT datasets, blood vessel segmentation and MRI brain tumor segmentation, Kvasir-SEG and CVC-ClinicDB datasets respectively.

Combining RS technology with the DL networks outperformed all previous methods for semantic segmentation in various domains, including agriculture. Incredibly, road and building extraction from high-resolution aerial images achieved remarkable performance through deep architectures based on UNet architecture. Deep residual UNet, proposed by Zhang et al. [20], replaces the plain convolutional blocks into residual blocks. The road extraction networks are improved with the extraction of fine details of the roads by He et al. [21] applying the atrous spatial pyramid pool (ASPP) in the encoder-decoder structure. Pre-trained encoders are employed instead of plain convolutional layers, and dilated convolutions at the decoder side are integrated into the D-LinkNet model proposed by Zhou et al. [22], which are suitable for extracting narrow, complex roads from high-resolution satellite images. UNet has also been modified to automatically detect the center pivot irrigation systems [23] and land cover mapping [24] from satellite images. To the best of our knowledge, no study has been reported yet on detecting the water-stressed sugarcane crop using semantic segmentation.

3. METHODOLOGY

Encoder-decoder architecture is the best approach among various approaches such as convolutional networks with graphical models, multiscale and pyramid network-based models, R-CNN-based models, dilated convolution, and DeepLab family, attention-based models, generative models, and adversarial training [25] for semantic image segmentation. Based on this encoder-decoder approach, after UNet, various research studies such as SegNet [16] and FCN [14] showed notable performances. The encoder designed in the UNet comprises two convolutional layers followed by the max pooling layer, which was enhanced to have pre-trained models such as VGG, ResNet that contains convolutional layer, batch normalization, ReLU activation, and max pooling layer. The decoder structure mainly contributes to the network's performance in segmentation which assigns discernible features to pixel space. Features extracted at the encoder path must be reused at the decoder path to avoid the loss of low-level details. Reusing low-level details and retaining high-level semantic information are the critical points in the success of semantic segmentation.

3.1 DenseResUNet

This paper proposes a novel neural network architecture that reuses the features generated at the encoding path. The backbone of this architecture is the classical UNet; however, it differs from it in many ways. Firstly, the traditional

convolutional layers are replaced with the DenseRes block that utilizes the power of deep residual networks [26] and densely connected networks [27]. Secondly, the reduction in features maps is performed using a down transition block instead of max pooling in UNet and restructured the decoder path with DenseRes block and Up transition block. In addition, the proposed network grasps the benefit of the attention module to focus on the essential features of the complex input. The model gains benefits of features reusability due to concatenating path and dense connection because of the dense block, increase in training speed owing to residual networks, and gradient degradation smooth information propagation in the network. The proposed DenseResUNet architecture consists of an encoder path, a bridge, and a decoder path with skip connections, as depicted in Figure 1. The encoding represents the input images in extracted features form, and decoding

performs a synthesis task to generate pixel-wise segmentation. The encoder path comprises four DenseRes blocks. Each of the DenseRes blocks includes five residual blocks that are densely connected, as shown in Figure 2. Each of the residual blocks contains three consecutive convolutional blocks. Every convolutional block applies a full pre-activation residual connection. A down transition block follows each DenseRes block to reduce the size of feature maps to half. The middle part of the network joins the encoder and decoder path with a single DenseRes block. The decoder path includes four DenseRes blocks with the same architecture as in the encoder's block. Each DenseRes block is connected to the attention module to focus on the vital area of the image. An up transition block follows the attention module to regain the input image resolution.

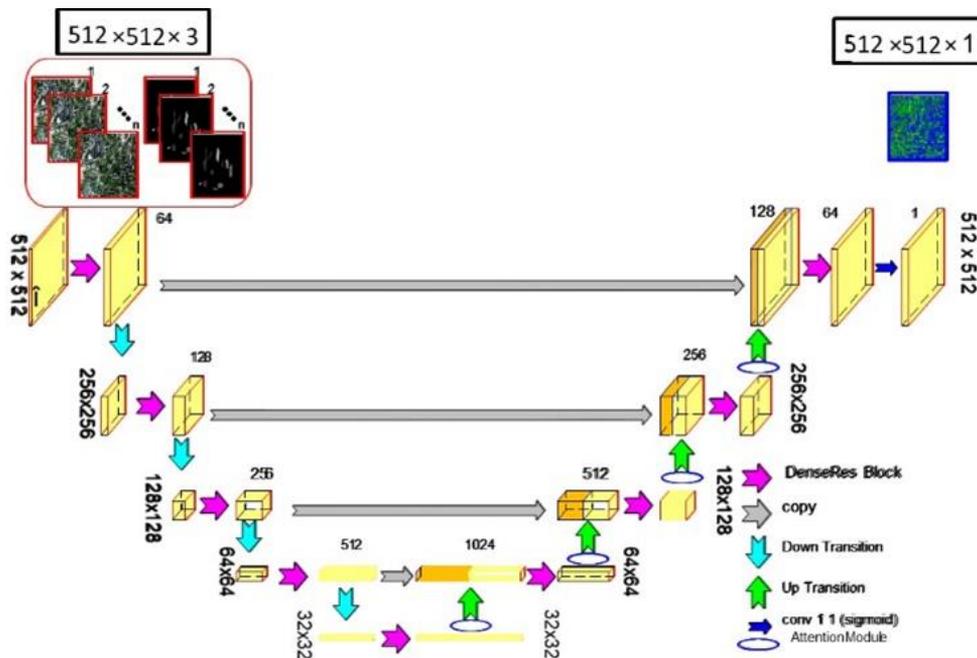


Figure 1. The proposed DenseResUNet architecture

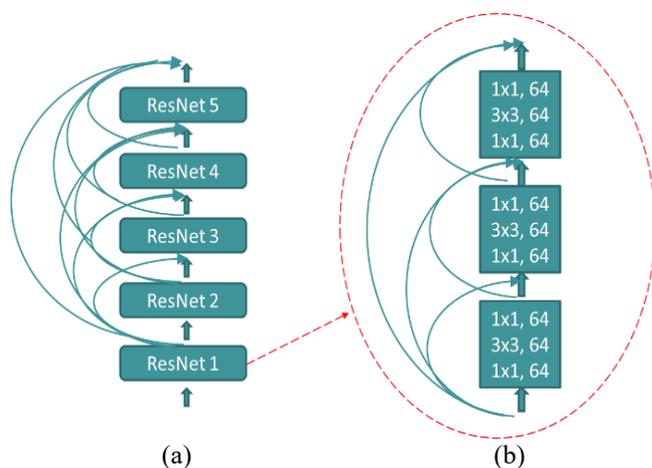


Figure 2. a) DenseRes block b) Layers in the residual block

A final 1x1 convolution layer with sigmoid activation is used at the end of the decoder path. This layer assigns the class to the pixels that generated a segmentation map. All the building blocks of the model are briefly explained in the following sections.

3.2 Residual block

The most common vanishing gradient problem in a deeper network was addressed by He et al. [26] with ease in the network's training with a deep residual neural network (ResNet). The ResNet comprises residual blocks with the fundamental idea of an identity mapping that allows information flow without compromising the performance of a deeper network. It also lessens the network parameters. Residual block adds up its output with the identity mapping is represented below:

$$Y_l = H_l(Y_{l-1}) + Y_{l-1} \quad (1)$$

where, $H_l(\cdot)$ is a residual function made up of convolutional layers, ReLU activation function, and batch normalization. The input of this l th residual block is Y_{l-1} , and the output generated by this block is Y_l . It has successfully experimented with semantic segmentation [20] and is the reason it is employed in the proposed architecture. The residual block uses two layers for the ResNet18, ResNet30 and three layers for ResNet50, ResNet101, and higher versions. The proposed architecture uses three layers in a residual block, as shown in

Figure 2(b). The reason behind choosing these residual layers is they do not increase the training error percentage and helping in easy network training. Residual units extract features with expedite training of the network. These units are embedded within the dense block of the proposed architecture Figure 2(a), which strengthens the features propagation across the network. Residual layers make training faster and dense block controls the vanishing gradient problem with fewer parameters.

3.3 DenseRes block

After ResNet's skip connection's immense success, another CNN architecture, DenseNet, densely connected network witnessed great success in computer vision tasks, primarily image classification tasks. The core of the DenseNet network is that each layer's feature maps are concatenated to every other layer in the network. These concatenations lead to easy network training and provide accurate results. These all feature maps are fed to a composite function, as shown below:

$$Y_l = H_l([Y_{l-1}, Y_{l-2}, \dots, Y_0]) \quad (2)$$

where, $H_l(.)$ is a composite function made up of batch normalization, ReLU, and a 3x3 convolution. The output of the l^{th} layer is which concatenates all previous layers' feature $[Y_{l-1}, Y_{l-2}, \dots, Y_0]$. A classical dense block contains five layers connected through the composite function. The idea of concatenating the feature maps generated of all previous layers to the next layer is borrowed in the proposed architecture. However, the composite function differs from the classical dense block by the replacement with a residual block (Figure 2a). The layers of DenseRes block and residual block are given in the following algorithm:

Algorithm 1: Algorithmic steps for DenseResBlock

```

Input: x-keras model network
Output: x- keras model network
1: List <- []
2: for i = 0 to NumberOfLayers do
3:   input <- x
4:   for i = 0 to 2 do
5:     InputToInnerBlock <- x
6:     x <- BatchNormalization(x)
7:     x <- ReLUActivation(x)
8:     x <- 2D convolution operation with 1x1 kernel(x)
9:     x <- BatchNormalization(x)
10:    x <- ReLUActivation(x)
11:    x <- 2D convolution operation with 3x3 kernel(x)
12:    if d then dropout
13:      x <- dropout(x)
14:    end if
15:    x <- BatchNormalization(x)
16:    x <- ReLUActivation(x)
17:    res <- 2D convolution operation with 1x1 kernel(x)
18:    IdentityMapping <- 2D convolution operation with 1x1kernel (InputToInnerBlock)
19:    IdentityMapping <- BatchNormalization(IdentityMapping)
20:    IdentityMapping <- ReLUActivation(IdentityMapping)
21:    x <- Add(res, IdentityMapping)
22:  end for
23: IdentityMappingOuter <- 2D convolution operation with

```

```

1x1 kernel(input)
24:IdentityMappingOuter <- BatchNormalization(IdentityMapping)
25:IdentityMappingOuter <- ReLUActivation(IdentityMapping)
26: x <- Add (x, IdentityMappingOuter)
27: ListAppend(x)
28: x <- Concatenate (List)
29: Update number of filters with growth rate
30: end for
31: return x

```

3.4 Attention mechanism

Attention mechanism (a.k.a. attention module) is an input processing unit that makes the network concentrate on specific network features at a time. That is, it works on the subset of the dataset. An attention mechanism requires continuous backpropagation to be effective in the network. The first attention mechanism was proposed by Bahdanau et al. [28]. It was an advancement in an encoder-decoder recurrent neural network / long short term memory (RNN/LSTMs) based neural machine translation system. Later, it was successfully applied in computer vision tasks [29]. Due to this property of the convolution block attention module and its remarkable performance in semantic segmentation, the attention module is integrated into the proposed architecture at the decoder path to focus on essential input aspects.

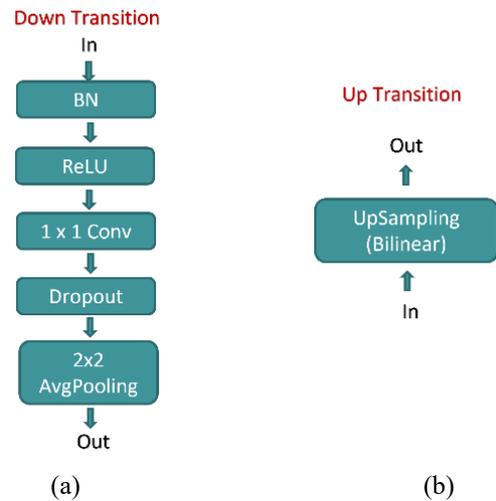


Figure 3. a) Down transition block b) Up transition block

3.5 Down and up transitions

The up transition and down transition blocks are described in Figure 3. The down sampling block consists of a batch normalization layer, ReLU activation, 1x1 convolution layer to retain the feature maps generated at the encoding end. Many complex connections among the layers refer to the noise in a multi-layer network if training data is limited. Referring to the noise leads to overfitting, wherein the model learns the details and noise to a great extent, adversely affecting the model's performance. The widely used successful technique to overcome the overfitting problem is called a dropout. Dropout is applied in the down transition block after the convolution in the proposed architecture as against the original UNet architecture. The dropout rate was set to a value of 0.25. Then, average pooling is used to bring down the size of the feature

maps at the time of encoding. Up transition block uses upsampling to bring up the size of the feature map during decoding, followed by a transpose convolution.

4. EXPERIMENTS

4.1 Dataset

A dataset is created containing water-stressed sugarcane crops from a region of four talukas (Gokak, Raibag, Jamkhandi, and Mudhol) in the Karnataka state of India with coordinates 16.3898° N and 75.0371° E. The region is as shown in Figure 4.

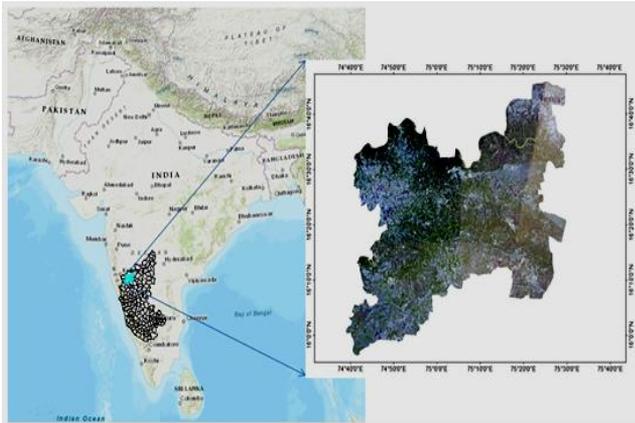


Figure 4. The study area

The region is dry with a temperature range between 16.20° C to 39.00° C. Among main crops such as banana, turmeric, and maize, sugarcane is cultivated in large quantities. A cloud-free Sentinel-2 satellite image captured on 27th May 2019 was downloaded from ESA's Copernicus website (www.scihub.Copernicus.eu). The downloaded image was pre-processed and analyzed for surface reflectance using the QGIS software program (Ver. 3.4.1). Red, near-infrared (NIR), and short-wave infrared (SWIR) bands of the imagery are utilized in this study (refer Table 1) as in near-infrared (~700 to 1300 nm) band leaves absorb less radiation and have high reflectance and transmission values, primarily influenced by leaf structural characteristics and biomass [30, 31]. Additionally, canopy reflectance patterns are influenced by the structure of the plant canopy and the area of the leaves [32, 33]. Water absorbs the most light in the short-wave infrared region (SWIR) between 1300 and 2500 nm, followed by other organic components present in leaves. The data of sugarcane field boundaries and other information such as crop growth stage, crop variety, date of planting, vigor, and GPS location (latitude, longitude) are collected from multiple locations. The water-stressed crop fields are marked manually with a handheld GPS device (Montana 680) through a field campaign conducted in May 2019. These shape files are rasterized and used to generate the corresponding ground truth images. The ground truth images are RGB channel images with labels of high-stress crops, medium stress crops, low-stress crops, and all other objects as background. The labeling is performed in LabelMe software. The dataset contains 400 surveyed polygons (shape files), and the rest of the labeling is prepared on this basis. The input Sentinel-2 image is split into patches of 512 x 512 pixels. These image patches are utilized for

training the proposed model to segment water-stressed sugarcane plots. 60% of images are used for training the network from total images, 25% are used for validation, and 15% are used for testing.

Table 1. Details of Sentinel-2 multispectral bands

Band number	Band Name	Spatial Resolution	Central wavelength (nm)
B4	Visible Red	10	665
B8	NIR	10	842
B11	SWIR	20	1610

4.2 Implementation details

4.2.1 Hyperparameters optimization

It requires training multiple models with the optimization algorithm, learning rate, training batch size, and epochs to select an optimal model. The achieved model in this study is implemented in the Tensorflow Keras deep learning framework. The kernels were initialized to HeUniform [34]. Adam was selected as an optimization algorithm after many experiments at the time of training. Among various algorithms such as SGD, RMSProp, and Adagrad, Adam proved superior because of faster convergence and accurate results. The beta parameters are kept same as $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-8}$. The learning rate was initially set to $1e-3$ with exponential decay of 0.996 after every epoch. The learning rate is scheduled during the training of the model. Training images are resized to 512x512. The model was converged in 50 epochs. The hardware and software configuration used to train the model includes Ubuntu operating system with an NVIDIA P100 GPU having 16GB memory, TensorFlow version 2.4.1 with Keras 2.4.3.

4.2.2 Data augmentation

Convolutional networks are data greedy; however, remote sensing lacks the labeled data. Generating sugarcane water-stressed samples is a very tedious, labor, and cost-intensive process. The generated dataset in the present study contains 378 images. Data augmentation technique was adopted to generate more training data from limited samples. It enhances the training dataset, induces implicit regularization [35], and tackles the unbalanced class problem. The training dataset is increased through data augmentation by applying six operations: horizontal flip, vertical flip, and rotation by 200, shifting horizontally with value 0.05, shifting vertically with value 0.05, and shear with range 0.05.

4.2.3 Evaluation metrics

The model's performance is quantitatively tested using five performance metrics: mean intersection over union (IoU), accuracy, recall, precision, and F1 score. The IoU is the proportion of the cover zone between reality and predicted area of interest on the ground to the territory encompassed by them, and mean IoU (mIoU) is computed by taking a mean of the IoU of each class. Accuracy gives predicted values that matches with actual values in percentage. Recall is given by the ratio of number of true positives with the number of true positives plus false negative. Precision is defined by the ratio of number of true positives with the number of true positives plus number of false positives. F1 score is defined as the harmonic mean of recall and precision.

5. RESULTS

DenseResUNet model is trained on the generated dataset of water-stressed sugarcane fields with three classes: high-stress, medium stress, and low-stress sugarcane crops. All other objects present in the surrounding area are considered as background. The proposed model achieved mIoU of 61.91% dataset. The proposed architecture is compared against three models UNet, ResUNet, and DenseUNet. The water-stressed sugarcane fields are segmented by UNet architecture with mIoU of 32.20%. The ResUNet gave mIoU 58.34% and DenseUNet achieved 53.15% mIoU. The other quantitative results in terms of accuracy, recall, precision, and F1 score on the testing dataset are presented in Table 2.

Table 2. Quantitative results of semantic segmentation models for water-stressed sugarcane crops dataset

Model	mIoU	Accuracy	Recall	Precision	F1
Unet	32.20	55.03	45.69	63.83	53.01
ResUNet	58.34	77.61	75.07	82.04	78.37
DenseUNet	53.15	72.65	69.78	76.93	73.15
DenseResUNet	61.91	80.53	77.98	84.53	81.11

Table 3. Comparison of time required to train the four models on the water-stressed sugarcane crops dataset

Model	Training Time (hrs)
Unet	2.10
ResUNet	2.14
DenseUNet	2.25
Proposed Model	2.50

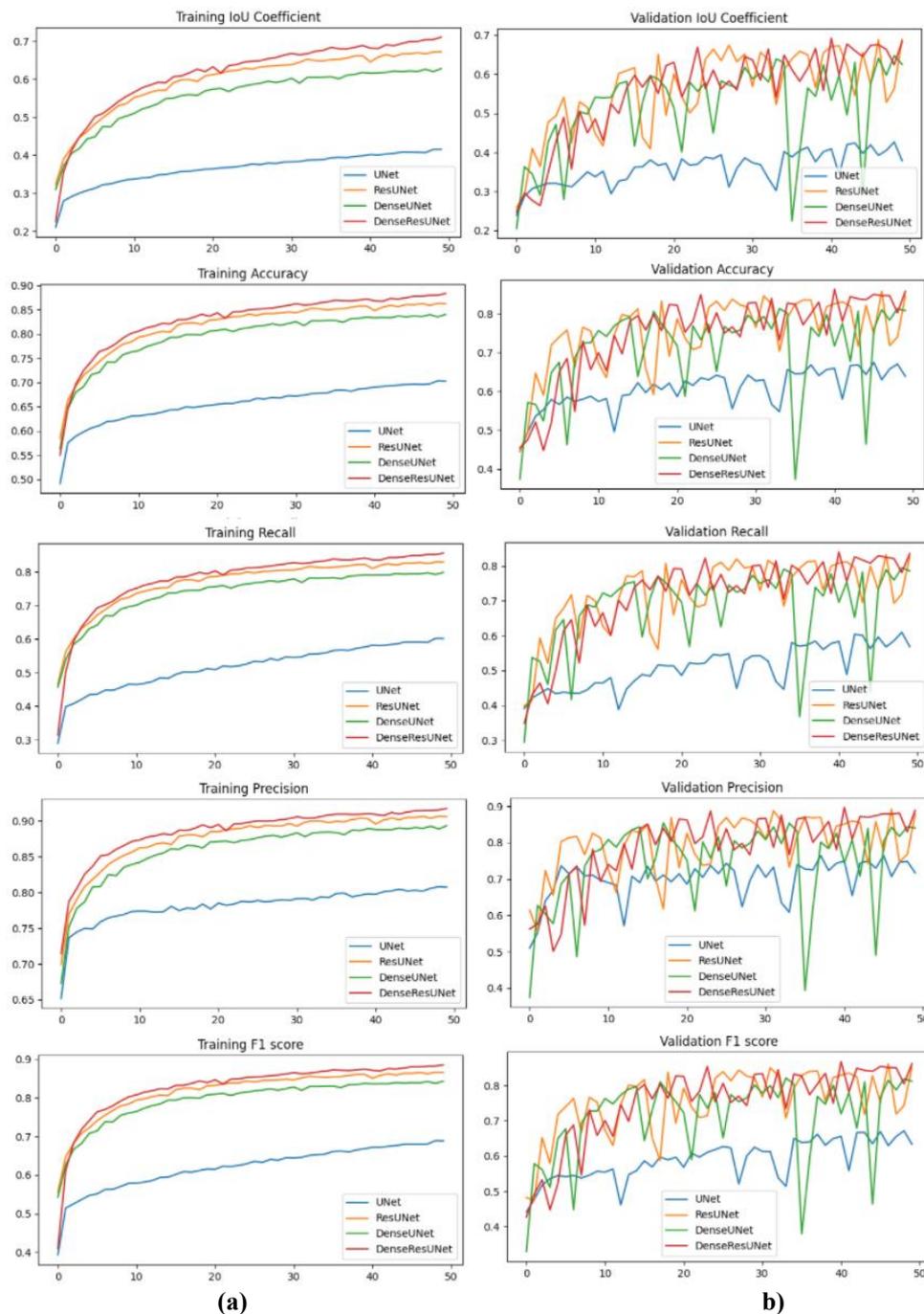


Figure 5. a) Training meanIoU, accuracy, recall, precision, and F1 score and b) validation meanIoU, accuracy, recall, precision and F1 score

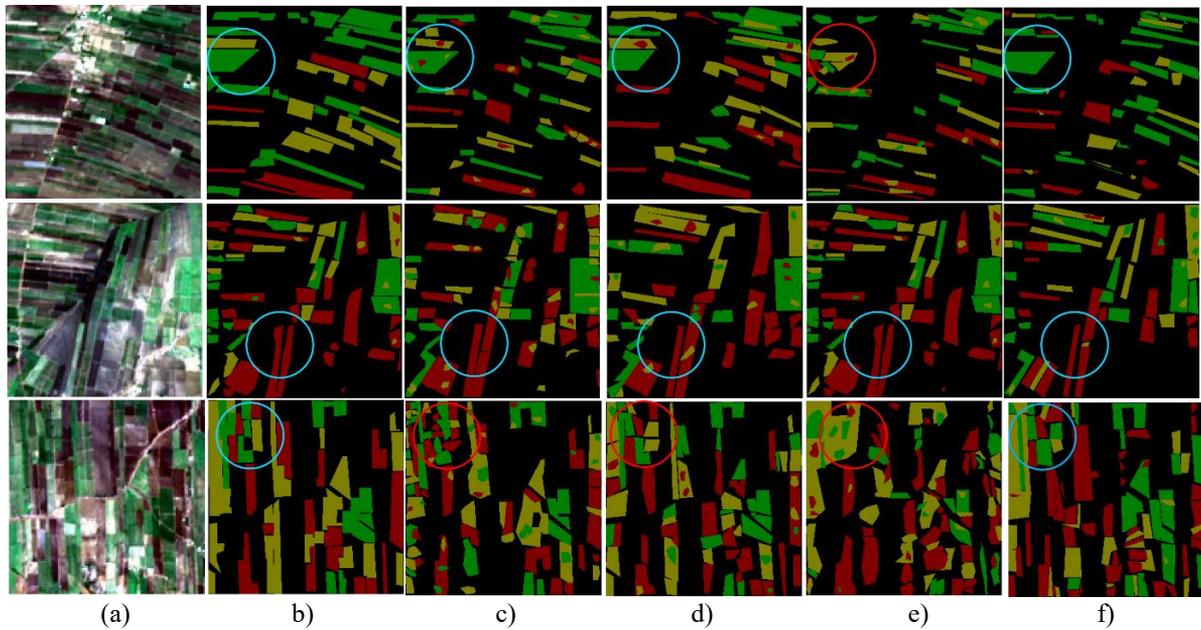


Figure 6. Example water-stressed sugarcane crops segmentation results through different models. a) The imagery of land b) Ground Truth image corresponding to the land imagery c) UNet model results d) ResUNet model results e) DenseUNet model results f) The proposed DenseResUNet model results. (Red color indicates high-stressed crops, green color shows medium-stressed crops, and yellow indicates low-stressed crops.)

The epoch-wise results of the four models are given in Figure 5. Time taken to train all the models is compared in Table 3. The DenseResUNet model took more time than all other models because of the many layers in it. The qualitative results are presented in Figure 6, wherein segmented images generated by all four models are compared. As observed from the results, the larger size plots with sugarcane crops are segmented with more accuracy than the plots with tiny size. Most of the high-stressed sugarcane crops and low-stressed crops are correctly classified. In medium stressed plots, models are confused and misclassified either into high-stress or into low-stress plots. For example, in the last row of Figure 6 (highlighted in a red circle and blue circle), various small plots are very close. In this case, UNet produced mixed results. ResUNet misclassified few low-stressed plots, whereas DenseUNet misclassified all high-stressed plots. Comparatively, DenseResUNet produced good results. Another example is shown in the first row of Figure 6 where DenseResUNet segmented low-stressed plots correctly achieved 80.53% accuracy on these water-stressed sugarcane crops. The results show the capability of semantic segmentation in crop water stress detection through limited remote sensing data. As we would increase the training samples, the results would be enhanced.

6. CONCLUSIONS

This paper presents a novel architecture, DenseResUNet, to segment water-stressed sugarcane crops from single date Sentinel-2 satellite imagery. The proposed architecture takes the benefits of dense connections, residual blocks, and an attention module. The dense connection aids accurate computation and skip connections between contracting and expansive paths, helping recover spatial information loss. In addition, features are reused in the architecture. The present study results clearly show the effectiveness of an attention module in an expansive path with dense blocks. The

adequately designed architecture can segment the sugarcane water-stressed plots with 61.91% mIoU and 80.53% accuracy. The architecture is compared with UNet, ResUNet, and DenseUNet models. The results of this study reveal its suitability in agriculture applications from remote sensing data.

REFERENCES

- [1] 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. *Precision Agriculture*, 21(5): 1121-1155. <https://doi.org/10.1007/s11119-020-09711-9>
- [2] Virnodkar, S.S., Pachghare, V.K., Patil, V.C., Jha, S.K. (2020). CaneSat dataset to leverage convolutional neural networks for sugarcane classification from Sentinel-2. *Journal of King Saud University-Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.09.005>
- [3] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [4] Virnodkar, S., Pachghare, V.K., Patil, V.C., Jha, S.K. (2021). Performance evaluation of RF and SVM for sugarcane classification using Sentinel-2 NDVI time-series. In *Progress in Advanced Computing and Intelligent Engineering*, pp. 163-174. https://doi.org/10.1007/978-981-15-6353-9_15
- [5] Zhou, B., Lapedriza, A., Xiao, J., Torrallba, A., Oliva, A. (2014). Learning deep features for scene recognition using places database. <http://hdl.handle.net/1721.1/96941>
- [6] Nemade, S.B., Sonavane, S.P. (2020). Co-occurrence patterns based fruit quality detection for hierarchical fruit image annotation. *Journal of King Saud University-Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2020.11.033>

- [7] Nemade, S., Sonavane, S. (2020). Stacked GRU based image Captioning. *Solid State Technology*, 63-2: 1052-1063.
- [8] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [9] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440. <https://arxiv.org/abs/1409.4842v1>.
- [10] Singh, A.K., Ganapathysubramanian, B., Sarkar, S., Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends in Plant Science*, 23(10): 883-898. <https://doi.org/10.1016/j.tplants.2018.07.004>
- [11] Chandel, N.S., Chakraborty, S.K., Rajwade, Y.A., Dubey, K., Tiwari, M.K., Jat, D. (2021). Identifying crop water stress using deep learning models. *Neural Computing and Applications*, 33(10): 5353-5367. <https://doi.org/10.1007/s00521-020-05325-4>
- [12] Gao, Z., Luo, Z., Zhang, W., Lv, Z., Xu, Y. (2020). Deep learning application in plant stress imaging: A review. *AgriEngineering*, 2(3): 430-446. <https://doi.org/10.3390/agriengineering2030029>
- [13] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.
- [14] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D. (2015). Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1529-1537. <https://doi.org/10.1109/ICCV.2015.179>
- [15] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 9351: 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [16] Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [17] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Transactions on Medical Imaging*, 37(12): 2663-2674. <https://doi.org/10.1109/TMI.2018.2845918>
- [18] Zhang, Z., Wu, C., Coleman, S., Kerr, D. (2020). DENSE-INception U-net for medical image segmentation. *Computer Methods and Programs in Biomedicine*, 192(105395): 1-15. <https://doi.org/10.1016/j.cmpb.2020.105395>
- [19] Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D. (2019). Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE International Symposium on Multimedia (ISM)*, pp. 225-2255. <https://arxiv.org/abs/1911.07067v1>
- [20] Zhang, Z., Liu, Q., Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 749-753. <https://doi.org/10.1109/LGRS.2018.2802944>
- [21] He, H., Yang, D., Wang, S., Wang, S., Li, Y. (2019). Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. *Remote Sensing*, 11(9): 1015-1031. <https://doi.org/10.3390/rs11091015>
- [22] Zhou, L., Zhang, C., Wu, M. (2018). D-linknet: Linknet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 182-186. <https://doi.org/10.1109/IGARSS.2019.8898392>
- [23] Saraiva, M., Protas, É., Salgado, M., Souza, C. (2020). Automatic mapping of center pivot irrigation systems from satellite images using deep learning. *Remote Sensing*, 12(3): 558. <https://doi.org/10.3390/rs12030558>
- [24] Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D. (2019). Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, 11(17): 1986. <https://doi.org/10.3390/rs11171986>
- [25] Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2021.3059968>
- [26] He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pp. 630-645. <https://arxiv.org/abs/1603.05027v3>
- [27] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700-4708. <https://arxiv.org/abs/1608.06993v5>
- [28] Bahdanau, D., Cho, K., Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- [29] Li, H., Xiong, P., An, J., Wang, L. (2018). Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*.
- [30] Avery, T.E., Berlin, G.L. (1992) *Fundamentals of Remote Sensing and Airphoto Interpretation*; Macmillan: London, UK.
- [31] Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2): 127-150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- [32] Rautiainen, M., Stenberg, P. (2005). Application of photon recollision probability in coniferous canopy reflectance simulations. *Remote Sensing of Environment*, 96(1): 98-107. <https://doi.org/10.1016/j.rse.2005.02.009>
- [33] Disney, M., Lewis, P., Saich, P. (2006). 3D modelling of forest canopy structure for remote sensing simulations in the optical and microwave domains. *Remote Sensing of Environment*, 100(1): 114-132. <https://doi.org/10.1016/j.rse.2005.10.003>
- [34] He, K., Zhang, X., Ren, S., Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on

ImageNet classification. In Proceedings of the IEEE International Conference on Computer Vision, pp. 1026-1034. <https://arxiv.org/abs/1502.01852v1>

[35] Simard, P.Y., Steinkraus, D., Platt, J.C. (2003). Best

practices for convolutional neural networks applied to visual document analysis. Seventh International Conference on Document Analysis and Recognition, pp. 958-963. <https://doi.org/10.1109/ICDAR.2003.1227801>