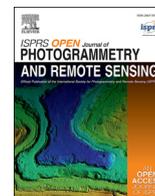


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

ISPRS Open Journal of Photogrammetry and Remote Sensing

journal homepage: www.editorialmanager.com/OPHOTO

Mapping sugarcane in Thailand using transfer learning, a lightweight convolutional neural network, NICFI high resolution satellite imagery and Google Earth Engine



Ate Poortinga^{a,b,*}, Nyein Soe Thwal^{b,c}, Nishanta Khanal^{a,b}, Timothy Mayer^{d,e}, Biplav Bhandari^{e,f}, Kel Markert^{d,e}, Andrea P. Nicolau^{d,e}, John Dilger^{a,b}, Karis Tenneson^{a,b}, Nicholas Clinton^g, David Saah^{a,b,h}

^a Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA, 94566, USA

^b SERVIR-Mekong, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok, 10400, Thailand

^c Asian Disaster Preparedness Center, SM Tower, 24th Floor, 979/69 Paholyothin Road, Samsen Nai Phayathai, Bangkok, 10400, Thailand

^d Earth System Science Center, The University of Alabama in Huntsville, 320 Sparkman Dr., Huntsville, AL, 35805, USA

^e SERVIR Science Coordination Office, NASA Marshall Space Flight Center, 320 Sparkman Dr., Huntsville, AL, 35805, USA

^f Department of Atmospheric and Earth Science, The University of Alabama in Huntsville, 320 Sparkman Dr., Huntsville, AL, 35805, USA

^g Google, Inc., 1600 Amphitheatre Parkway, Mountain View, CA, 94043, USA

^h Geospatial Analysis Lab, University of San Francisco, 2130 Fulton St., San Francisco, CA, 94117, USA

ARTICLE INFO

Keywords:

Mekong region
High resolution satellite imagery
MobileNetV2
U-net convolutional network
Deep-learning
Artificial intelligence

ABSTRACT

Air pollution from burning sugarcane is an important environmental issue in Thailand. Knowing the location and extent of sugarcane plantations would help in formulating effective strategies to reduce burning. High resolution satellite imagery combined with deep-learning technologies can be effective to map sugarcane with high precision. However, land cover mapping using high resolution data and computationally intensive deep-learning networks can be computationally costly. In this study, we used high resolution satellite imagery from Planet that has been made available to the public through the Norway's International Climate and Forest Initiative (NICFI). We tested a U-Net deep-learning algorithm with a lightweight MobileNetV2 network as the encoder branch using the Google Earth Engine computational platform. We trained a model using the RGB channels with pre-trained network (RGBt), a RGB model with randomly initialized weights (RGBr) and a model with randomly initialized weights including the NIR channel (RGBN). We found an F1-score of 0.9550, 0.9262 and 0.9297 for the RGBt, RGBr and RGBN models, respectively. For an independent model evaluation we found F1-scores of 0.9141, 0.8681 and 0.8911. We also found a discrepancy in the recall values reported by the model and those from the independent validation. We found that lightweight deep-learning models produce satisfactory results while providing effective means to apply mapping efforts at scale with reduced computational costs. We highlight the importance of central data repositories with labeled data as pre-trained networks were found to be effective in improving the accuracy.

1. Introduction

Agricultural and industrial development processes have been a major driver behind land cover changes in Southeast Asia (FAO, 2020; FAO, 2019; Poortinga et al., 2020). Whereas crop yields have soared over the last two decades and dramatically improved the food security situation (Poortinga et al., 2019a), agricultural expansion into territory with high value in terms of ecology and biodiversity remains a main concern.

Moreover, agricultural production systems are a large source of greenhouse gas emissions due to the use of fertilizer, fossil fuels and biomass burning (Yuttitham et al., 2011). Biomass burning is also a main source of hazardous air pollution and associated negative health consequences (Chantara et al., 2012).

Thailand, like many other countries in Southeast Asia is a major exporter of a number of agricultural food and industrial products. Main crop commodities in Thailand include rice, soybean, cassava, and

* Corresponding author. Spatial Informatics Group, LLC, 2529 Yolanda Ct., Pleasanton, CA, 94566, USA.

E-mail address: apoortinga@sig-gis.com (A. Poortinga).

<https://doi.org/10.1016/j.ophoto.2021.100003>

Received 18 January 2021; Received in revised form 29 June 2021; Accepted 19 July 2021

Available online 24 July 2021

2667-3932/© 2021 The Author(s). Published by Elsevier B.V. on behalf of International Society of Photogrammetry and Remote Sensing (isprs). This is an open access

article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

sugarcane (Tenneson et al., 2021). In Thailand, agricultural production is not a main driver for deforestation like other countries in Southeast Asia (Tenneson et al., 2021). However, agricultural production is a primary concern for environmental impact including, but not limited to, issues regarding water (quality and quantity) (Silalertruksa and Gheewala, 2018), soil health (Oechaiyaphum et al., 2020) and air pollution (Vadrevu et al., 2015). Additionally, the sustainability of the agricultural production system in relation to human health is a complex interaction that includes socio-economic, geographical and climatological factors (Walters et al., 2016). Detailed geographically explicit information is required to map and understand how those multidimensional components interact.

Maps depicting the location, coverage, and growth of specific crop commodities are critical to understand the drivers behind agricultural practices and could help formulate strategies to improve sustainability of agro-ecological practices. Whereas many countries lack accurate and timely land cover data (Saahet al.), Thailand has a precise archive of spatially explicit land cover information as well as census data. However, the hand digitized land cover data provides a single snapshot and does not represent spatio-temporal dynamics in cropping patterns. The work of Saah et al. (2020), Potapov et al. (2019) provides a comprehensive overview of yearly land cover change dynamics for the period 2000–2018 using machine learning, however, their analyses were done using medium resolution Landsat data and their typology does not include specific crop commodities. Data with a finer spatio-temporal resolution would help circumvent current issues in land cover mapping including persistent cloud cover (Saahet al.) and mixed pixels. Furthermore, deep-learning technologies have demonstrated to improve land cover mapping efforts (Zhang et al., 2020).

High resolution images from both drones and satellites have limitations on the spatial and temporal coverage, historical archive, cost, and weather obstructions such as cloud cover or inability to fly due to inclement weather (Bhandari et al., 2015; Fisher et al., 2018). Historically, due to the financial overhead from both storage and computational demands, high resolution satellite imagery for land cover mapping has been beyond the reach of most organizations. However, recently, high resolution imagery has been made available to the public through the Norway's International Climate and Forest Initiative (NICFI), with Kongsberg Satellite Services (KSAT) and its partners Airbus and Planet. Moreover, cloud storage and computational frameworks such as Google Earth Engine have removed barriers with regards to computational and storage requirements (Gorelick et al., 2017). The recent integration of the Google Artificial Intelligence (AI) platform with Google Earth Engine enables scientists to apply deep-learning algorithms to geographic data at scale. The integration of cloud computing, high resolution satellite imagery and deep-learning technologies provides scientists with the tools to map the Earth's surface with ever increasing precision and revolutionizes our scientific understanding of the Earth.

Mapping crop commodities using satellite imagery is a well established field. There is a large scientific body of literature describing the use of publicly available data from optical sensors such as Moderate Resolution Imaging Spectroradiometer (MODIS) [e.g. Friedl et al., 2002], Landsat [e.g. Khanal et al., 2020; Uddin et al., 2015] and Sentinel-2 [e.g. Poortinga et al., 2019b; Tiwari et al., 2020] to map a wide variety of land cover types. Optical remote sensing methods mainly rely on the spectral signature of the crop type, i.e. the part of the solar spectrum reflected by the land cover type under investigation. Land cover phenology can add additional contextual information as some crop types show distinct temporal patterns in their spectral reflectance. In addition to conventional red, green, and blue channels, publicly funded medium resolution satellites additionally have multiple bands in the near infrared (NIR) and shortwave infrared (SWIR) part of the spectrum, which most high resolution imagery lack. Including these NIR and SWIR bands, as well as derived spectral indices, are often an advantage when using machine learning approaches. However, the PlanetScope Dove satellite series utilized for this study, are equipped with spectral bands including Blue

(455–515 nm), Green (500–590 nm), Red (590–670 nm), and NIR (780–860 nm) (Lemajic et al., 2018). The spectral ranges of these bands are comparatively larger than that of Landsat's complementary bands (Young et al., 2017). In total, the spectral range of Landsat, with the additional bands, surpasses the PlanetScope Dove. However, deep-learning technologies that take into account spatial information (e.g. convolutions) combined with high resolution data can help overcome limited collective spectral range compared to that of Landsat.

Traditional methods of land cover mapping include the use of machine learning approaches such as Vector Quantization (Gray, 1984), Random forest (Breiman, 2001) and Support Vector Machine (SVM) (Noble, 2006). Deep-learning technologies such as Deep Neural Networks (DNN), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have become increasingly popular in remote sensing. Work of e.g. Hoese and Kuenzer (Hoese and Kuenzer, 2020; Hoese and Kuenzer, 2020), Kattenborn et al. (2021), Pashaie et al. (2020), Paoletti et al. (2019) provide comprehensive overviews of recent trends and applications in the field of Earth observation. The technologies are based on the interdisciplinary field of computer vision and are applied on an operational basis in many scientific disciplines as well as commercially. Large open databases with imagery (Deng et al., 2009) are available for neural networks which are known to be much more data demanding than conventional machine learning approaches. Moreover, a variety of well documented and tested model architectures with pre-trained networks are available. Examples include the VGG-19 (Simonyan and Zisserman, 2014), Xception (Chollet, 2017) and MobileNetV2 (Sandler et al., 2018) integrated within the Python Keras deep-learning Application Programming Interface (API). Pre-trained models are highly effective for model-based transfer learning (Pan and Yang, 2009; Marmanis et al., 2015), but they only include the RGB channel and do not support the near infrared and shortwave infrared channels commonly available in satellite imagery (Panboonyuen et al., 2019). However, most pre-trained models contain a large number of variables and are computationally demanding during training and inference. Lightweight models would yield large cost reductions when applied at scale using high resolution satellite imagery for field scale mapping purposes.

The objective of this study is to apply cloud enabled deep-learning technologies at scale using high resolution satellite imagery to map sugarcane areas in support of reducing agricultural burning. We use three different strategies to map sugarcane in Thailand with a lightweight MobileNetV2 network. We tested the network using (1) the RGB channels in a pre-trained network (referred to as RGBt hereafter), (2) the RGB channels with randomly initialized weights (RGBr) and (3) the RGB and Near InfraRed channels is modeled with randomly initialized weights (RGBN). We report and compare different performance indicators from the model and from an independent validation data set. Innovations of this study include the application of deep-learning at scale using high resolution satellite imagery in Google Earth Engine. The results of this study should help guide the scientific community in applying effective lightweight deep-learning networks using high resolution satellite imagery for environmental mapping.

2. Materials and methods

2.1. Study region

This study was conducted for Thailand (Fig. 1). Thailand is located in Southeast Asia and has a population of approximately 70 million. The agricultural sector was transformed from a system that predominantly relied on subsistence farming system in the early 1960's to a system focused on commercial crops in the 1980's (Trisurat et al., 2019). Whereas agricultural exports are still an important economic activity in Thailand, the manufacturing and service sectors have been the main focus for economic growth. However, the expansion and intensification of the agricultural production system has also taken its toll on the environment (Santiphop et al., 2012).



Fig. 1. Thailand is located in Southeast Asia.

2.2. Workflow

The workflow is shown in Fig. 2. The NICFI Planet basemaps were uploaded to a Google Cloud Storage bucket and then ingested into Google Earth Engine as assets. The data was then sampled using annotated

sugarcane maps from the Land Development Department of Thailand. The data samples were exported as TensorFlow Records, aka TFRecords, to the cloud storage bucket. A Virtual Machine was then used to train the TensorFlow model using those TFRecords. The trained model was then deployed to the Google AI platform. Finally, the deployed model was used in the Earth Engine to perform inference for the country of Thailand.

2.3. Data description

The Planet surface reflectance basemaps that were made available through the NICFI were used in this study. This data has a spatial resolution of 4.77 m. We used the 2017 imagery that covered the period from June–November. Planet basemaps are particularly suitable for computer vision analytics as they are pre-processed data that account for sensor characteristics, sun angle, spatial accuracy and other artifacts caused by haze, light and topography. These basemaps were created from the PlanetScope satellite constellation. The 4-band PlanetScope imagery was corrected using the 6SV (Second Simulation of the Satellite Signal in the Solar Spectrum-Vector version (Vermote et al., 2006)) in combination with MODIS data while accounting for sun angle and satellite view geometry.

2.4. Data sampling

Hand digitized polygons of sugarcane from the Land Development Department of Thailand were used to train the model. However, this data was created in 2015 whereas, the satellite imagery used was taken from 2017. Due to this temporal misalignment between the satellite imagery and the polygons, data was filtered to areas that remained cropland over the two year period, and these locations were included for sampling. The polygons were rasterized to match the spatial resolution of the satellite imagery. A total of 100,511 data points were randomly placed across those areas. The data was subset into three portions, 70 % was used for training, 20 % for testing and 10 % for validation. Image patches were created at the center of each randomly distributed point. These image patches were 256 by 256 pixels and were then exported with binary information on sugarcane and the red, blue, green, and NIR bands.

2.5. Model architecture

We used a U-Net convolutional network (Fig. 3 (Ronneberger et al., 2015)) with a MobileNetV2 in the contraction path (encoder).

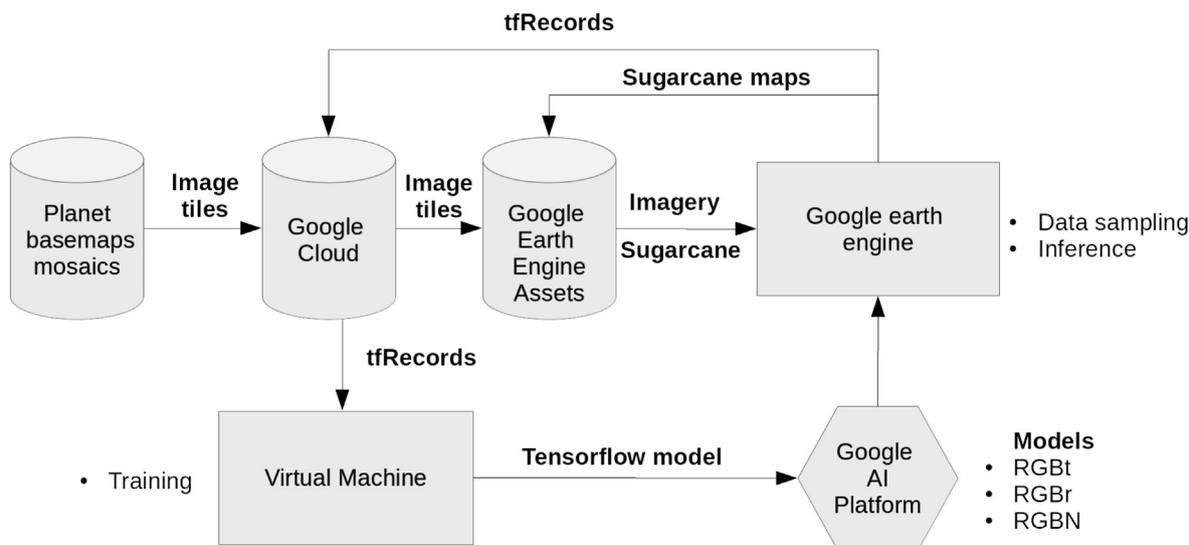


Fig. 2. The workflow includes the Google cloud bucket for data storage, Google Earth Engine Asset storage, the Earth Engine compute platform for data sampling and inference, a virtual machine for training and the Google AI platform where the models were deployed.

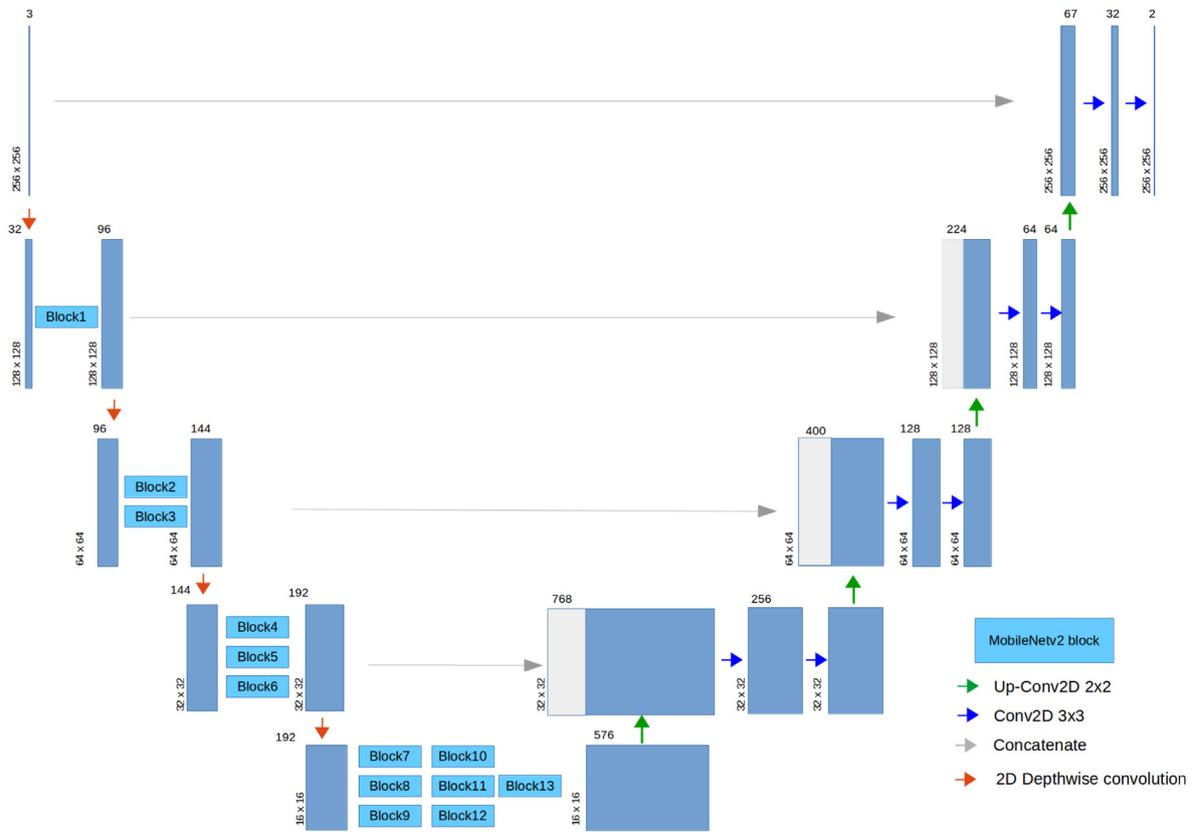


Fig. 3. U-Net model architecture used in this study. The MobileNetV2 was used as the encoder branch.

MobileNetV2 is an improvement on the MobileNetV1 (Howard et al., 2017) which is a lightweight network developed to efficiently balance performance and accuracy. The main innovation from MobileNetV1 was to replace convolutional layers with depth-wise separable convolutions which are computationally much cheaper. The convolutional layers were replaced by a 3×3 depth-wise convolution followed by a 1×1 point-wise convolution layer. Whereas the point-wise convolution layer kept the number of dimensions the same or made them larger in version 1, it is used as a ‘bottleneck’ or ‘projection layer’ in version 2, which means that it reduces the number of dimensions. Moreover, an expansion layer was added at the beginning of the block and a residual connection was added.

The first 13 blocks of the MobileNetV2 were included in the U-Net convolutional network (Fig. 3). In the encoder, convolutions instead of max pooling operation are used to reduce the spatial dimensions. The number of channels are denoted on top of each block. For example, the lowest resolution is 16×16 pixels and contains 576 channels. The input image, block 1, 3, 6 and 13 are connected to the decoder and 2×2 deconvolutions are used for upsampling. The decoder includes the conventional 3×3 ReLU convolutions in the upsampling path. We used a sigmoid activation function at the output layer to represent pixel sugarcane probability. Furthermore, we used the Adam optimizer (Kingma and BaAdam, 2014) with default settings. Data augmentation (Shorten and Khoshgoftaar, 2019) was applied by randomly flipping and rotating the data patches. Data augmentation operations were randomly applied to 70 % of the input data.

2.6. Performance evaluation

The accuracy (eq. (1)) and F1-score (eq. (2)) were used as the metric for model performance in this study. The F1-score is calculated using the precision and recall. The precision (eq. (4)) is the ratio of correctly predicted positive observations to the total predicted positive observations.

The recall (eq. (3)), also referred to as sensitivity, represents the ratio of correctly predicted positive observations to all the observations in the class. The F1-score (eq. (2)) is a weighted average of precision and recall. It takes into account both the false positives and false negatives.

$$accuracy = \frac{TN + TP}{TP + FP + TN + FN} \quad (1)$$

$$F1 = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (2)$$

$$recall = \frac{TP}{TP + FN} \quad (3)$$

$$precision = \frac{TP}{TP + FP} \quad (4)$$

where:

TP is the True Positives, which means that the actual class and the predicted class are both positive.

TN is the True Negatives, which means that the actual and predicted class are both negative.

FP is the False Positives, which means that the actual class is negative whereas the predicted class is positive.

FN is the False Negative, which means that the actual class is positive but the predicted class is negative.

Besides model validation, a manual independent validation was conducted. A random stratified sampling approach was used for each of the three model outputs and the combined outputs. Firstly, 100 random stratified points were placed on areas with a probability higher than 50 % and 167 on areas with a probability lower than 50 % on each map. Furthermore, 50 additional points were placed on areas with probabilities higher than 50 % on all three maps as well as for all combinations of 2 maps. This resulted in a total of 1001 points with half the points

associated with sugarcane plantations and the other half with other land cover classes. The points were shuffled and uploaded to Collect Earth Online (CEO (Saah et al., 2019)). These were classified using the 2017 NICFI imagery in CEO.

3. Results

The training and testing performance of the RGBt, RGBr and RGBN models are shown in Fig. 4 (top). It can be seen that the model using model-based transfer learning has the best and smoothest performance in terms of loss function, F1-score, precision, and recall. The RGBr and RGBN models have a comparable performance in terms of training, but the loss function and F1-score of the RGBN model shows more erratic behaviour between the iterations. The RGBr model shows more aberrant behaviour during testing in the early stage of the training.

The model validation results are shown in Table 1. It can be seen that the F1, precision, and recall score are the same for each model. The highest model accuracy was found for the RGBt model, followed by the RGBr and the RGBN model. The best and worst model performance is consistent for all measurements, the F1-score, precision, and recall. Inference was run on Google Earth Engine for all three models. We show the results for the Lop Buri province in Fig. 5. The three different models are shown and the sum of all models (bottom right), where a 3 means that all models predict sugarcane and 0 means that 0 models predict sugarcane. While there is much agreement for areas with dense sugarcane plantations, we find disagreement around the edges. The RGBt model covers a generally larger area than the other models.

Inference was ran for all three maps to calculate sugarcane pixel probability from the high resolution satellite imagery. Furthermore, a total of 1001 sample points were manually classified in CEO. The data points were used to sample the probability distributions of the three maps (Fig. 6). It was found that pixels classified as other have a distribution close to 0 for all models, but outliers exist. For sugarcane, it can be seen that the RGBt model has the highest median (93 %), followed by the RGBN (79 %) model and the RGBr model (74 %). The RGBt model has the highest minimum (34 %), again followed by the RGBN model (10 %) and the RGBr model (0 %). Fig. 6 displays the RGBt model having the least

Table 1

The accuracy, F1-score, precision and recall for the RGBt, RGBr and RGBN models.

	Accuracy	F1	Precision	Recall
RGBt	0.9554	0.9550	0.9550	0.9550
RGBr	0.9265	0.9262	0.9262	0.9262
RGBN	0.9003	0.9297	0.9297	0.9297

overlap between the two classes, Sugarcane and other, while the RGBr model displayed the most overlap for the classes. This is most evident with the long whiskers associated with the RGBr sugarcane class.

We used a 50 % threshold to convert the probability map into a binary map for sugarcane and other. The accuracies, F1-scores, precision and recall of the independent validation points are shown in Table 2. It was found that the RGBt model has the highest accuracy, followed by the RGBN and RGBr. The same order applies to the F1 scores and recall. However, the RGBr model has the best performance in terms of precision. For all models, the precision is higher than the recall. It can also be noted that the independent validation accuracies vary from the ones reported by the model. Differences in recall between the model and independent validation are most notable.

4. Discussion

In this study we used a MobileNetV2 U-Net convolutional network for semantic segmentation of sugarcane in Thailand. For three different approaches we found model accuracies between 90 and 95 %, but an independent assessment resulted in accuracies between 86 % and 91 %. These accuracies are in line with previously reported numbers. For example, Ma et al. (2019) reported a median accuracy of around 92 % for deep-learning approaches in land cover mapping classification studies. Yosinski et al. (2014) found that transfer learning yields better results than randomly initialized weights. These findings are supported by this study where the transfer learning model had the best performance. Pires de Lima and Marfurt (Pires de Lima and Marfurt, 2020) applied transfer learning for sensing scene classification and state that transfer learning is a powerful tool for remote-sensing scene classification. Moreover, they

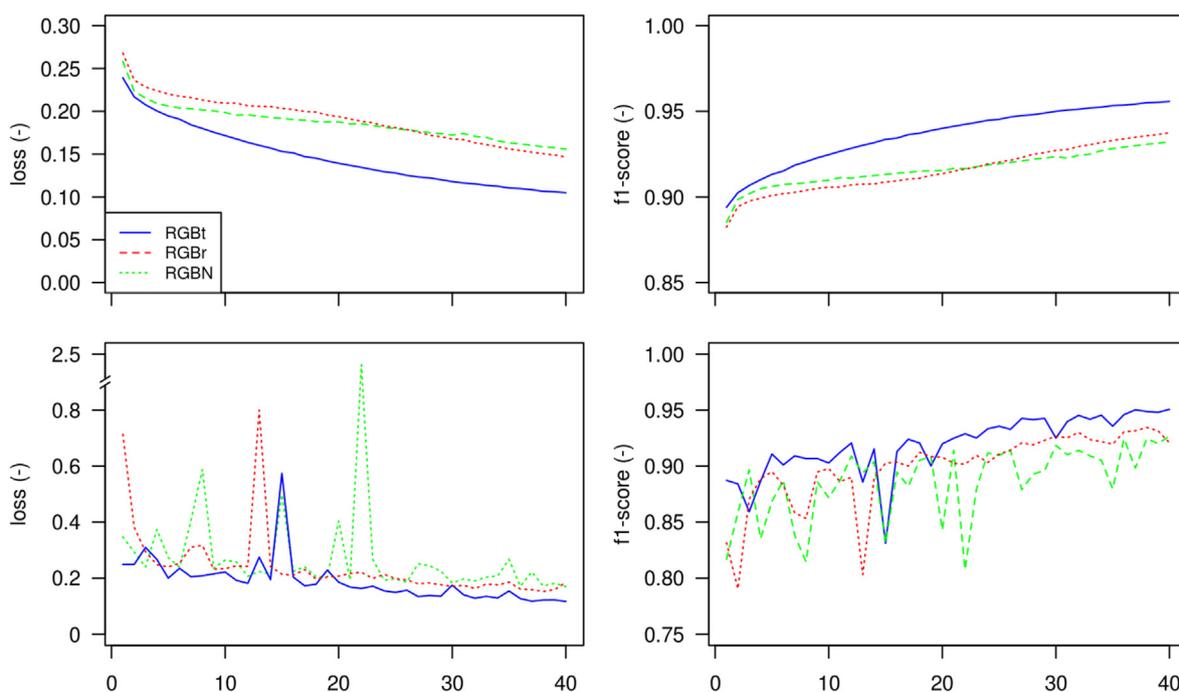


Fig. 4. Loss (left) and F1-score (right) for training (top) and testing (bottom). The different colors represent training using only RGB channels with transfer learning (blue), without transfer learning (red) and without transfer learning including the NIR channel (green).

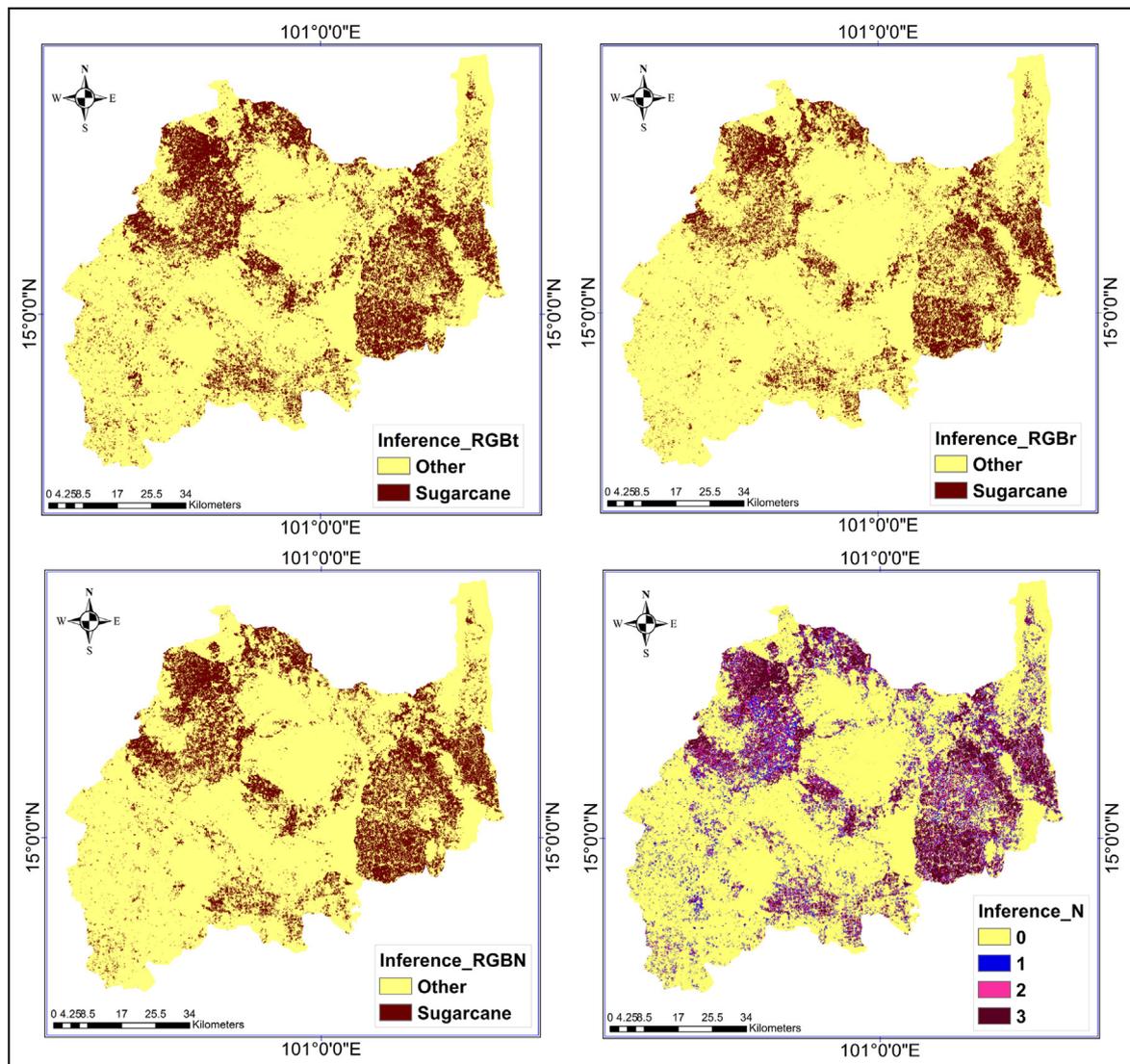


Fig. 5. Results of inference for Lop Buri for the RGBt (top left), RGBr model (top right) RGBN model (bottom left) and the combination of all models where the number represents the number of model agreements.

note that randomly initialized weights are also an appropriate choice for training given that the datasets are large enough. They used 21,800 samples whereas around 70,000 training samples were used in this study. The improved performance of the pre-trained network highlights the importance of large scale remote sensing archives such as BigEarthNet (Sumbul et al., 2019) and Radiant MLhub to develop and share pre-trained networks that cover multiple parts of the electromagnetic spectrum using various observation technologies. Dimensional reduction could be another viable strategy to reduce the number of spectral bands (Haut et al., 2018).

The independent sampling strategy highlights the discrepancies between the performance statistics reported by the model and the ones found using an independent sample. The most notable difference is the difference in recall, which was found to be much lower in the independent validation. This difference indicates a higher number of false negatives, which is specifically pronounced in the RGBr model. The precision scores are rather high, which indicates a low number of false positives in the data. The discrepancy between the two might be caused by spatial auto-correlation due to overlap in the patches. It should also be noted that the independent sample was much smaller than the modeled one. Nonetheless, we recommend to include an independent validation when using deep-learning techniques for remote sensing and Earth science

applications. Independent accuracy estimates can also be used for area estimations and error quantification (Olofsson et al., 2013, 2014).

The integration of Google AI platform with Google Earth Engine enables deploying deep-learning technologies at scale. The use of lightweight networks can be beneficial as they are computationally much more efficient and reduce costs. However, training and inference of large networks is computationally demanding. The models in this study were trained using 4 NVIDIA TESLA K80 GPU's, 24 cpu's and 224 GB RAM, whereas inference was done on the cloud using the Google AI GEE integration. Computational demands intensify when applying these networks on high resolution satellite imagery at scale. The mobileNet networks were developed for their computational efficiency, due to the limited number of parameters and computational efficiency. The study of Yu et al. (2019) used a MobileNetV2 Network for ship detection and reported a five times speed enhancement compared with conventional methods.

A limitation of the current modelling approach is that the model was trained using a remote sensing composite containing data between June and November. As such, no phenological characteristics were included in the model. Methods to include phenology-based features could include using multi-temporal composites where all bands are added as layers in the network. However, a different model design would be required to

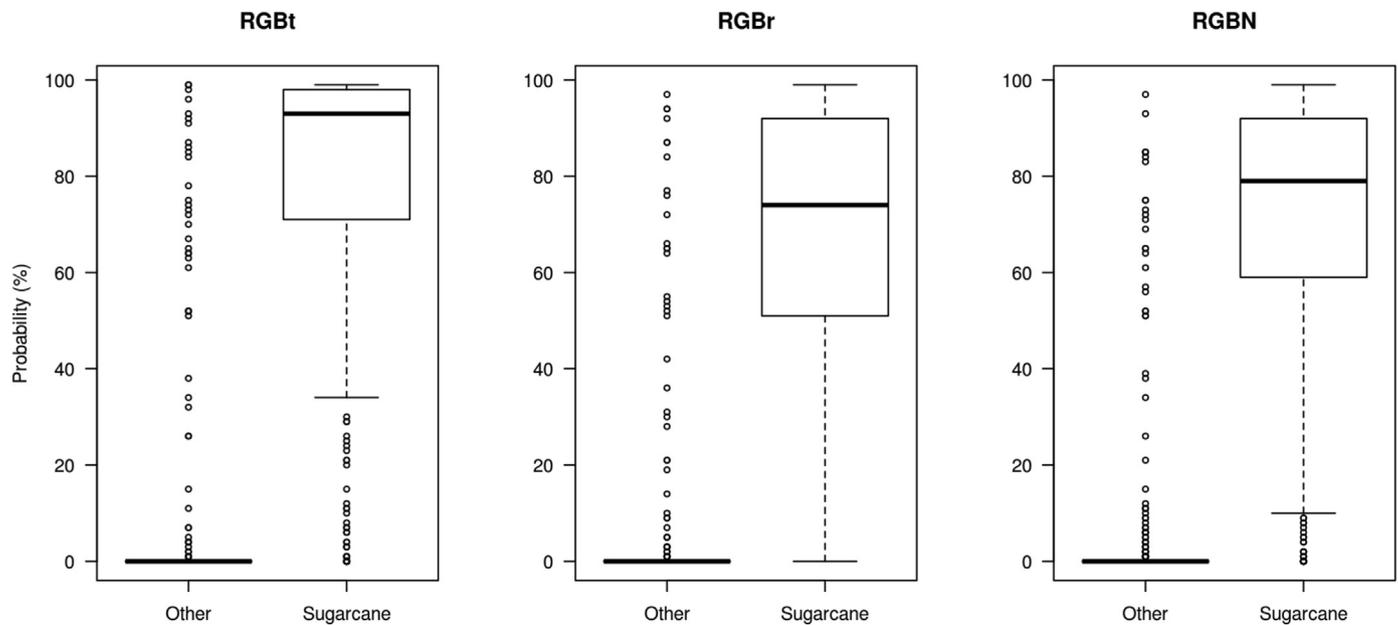


Fig. 6. Probability distribution of independent validation (1001 points) for the rgb model with transfer learning (RGBt), the rgb model without transfer learning (RGBr) and the RGBN model without transfer learning (RGBN).

Table 2
Independent validation results.

	Accuracy	F1	Precision	Recall
RGBt	0.9141	0.921	0.949	0.895
RGBr	0.8681	0.885	0.964	0.819
RGBN	0.8911	0.903	0.957	0.855

allow for transfer learning using composites from multiple dates, as demonstrated e.g. by the work of Zhao et al. (2020). They combined transfer learning with phenological information using a decision tree. Recurrent Neural Networks are also often used to map phenological characteristics. For example, (Crisóstomo de Castro Filho et al., 2020) used a long short-term memory (LSTM (Hochreiter and Schmidhuber, 1997)) to map rice crop, whereas Sun et al. (2019) implemented a workflow using Google Earth Engine to predict Soybean Yield with a CNN-LSTM Model.

In this study we demonstrated that the integration of cloud computing technologies, AI and open and free high resolution satellite imagery enables us to map the Earth's surface with great precision. Understanding the location, frequency of burning and other physical and socio-economic drivers will help formulate effective strategies to reduce burning, improve air quality and reduce health risks for millions of people. The data generated in this study will be combined with other data-sources on historical fire patterns and other socio-economic data.

5. Conclusion

We conclude that: (1) lightweight deep-learning models can be effective strategies to map land cover within large areas with accuracies greater than 86 %; (2) model-based transfer learning can be an effective strategy for land cover mapping, especially when limited data is available and furthermore yield higher accuracies than randomly initialized networks; (3) an independent validation using a robust sampling strategy is advised to evaluate final model performance.

Funding

This research was funded by the joint US Agency for International Development (USAID) and National Aeronautics and Space

Administration (NASA) initiative SERVIR-Mekong, Cooperative Agreement Number: AID-486-A-14-00002. Individuals affiliated with the University of Alabama in Huntsville (UAH) are funded through the NASA Applied Sciences Capacity Building Program, NASA Cooperative Agreement: NNM11AA01A.

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.

Acknowledgment

We thank the Land Development Department of Thailand for their generous support in providing land cover information that was used as the label data-set in this study. We thank Google for their generous support with providing free cloud credits that made the work possible. Our appreciation goes to the four anonymous reviewers and editor for their constructive comments and feedback that improved the quality of the manuscript.

References

- Bhandari, B., Oli, U., Pudasaini, U., Panta, N., 2015. Generation of high resolution dsm using uav images. In: FIG Working Week, pp. 17–21.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Chantara, S., Sillapapiromsuk, S., Wiriya, W., 2012. Atmospheric pollutants in chiang mai (Thailand) over a five-year period (2005–2009), their possible sources and relation to air mass movement. *Atmos. Environ.* 60, 88–98.
- Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258.
- Crisóstomo de Castro Filho, H., Abílio de Carvalho Júnior, O., Ferreira de Carvalho, O.L., Pozzobon de Bem, P., dos Santos de Moura, R., Olino de Albuquerque, A., Rosa Silva, C., Guimarães Ferreira, P.H., Fontes Guimarães, R., Trancoso Gomes, R.A., 2020. Rice crop detection using lstm, bi-lstm, and machine learning models from sentinel-1 time series. *Rem. Sens.* 12, 2655.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
- FAO, 2019. Forest Futures: Sustainable Pathways for Forests, Landscapes and People in the Asia-Pacific Region: Asia-Pacific Forest Secor Outlook Study III. Technical Report.
- FAO, U., 2020. The State of the World's Forests 2020: Forests, Biodiversity and People.

- Fisher, J.R., Acosta, E.A., Dennedy-Frank, P.J., Kroeger, T., Boucher, T.M., 2018. Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Rem. Sens. Ecol. Conserv.* 4, 137–149.
- Friedl, M.A., McIver, D.K., Hodges, J.C., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., et al., 2002. Global land cover mapping from modis: algorithms and early results. *Rem. Sens. Environ.* 83, 287–302.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Rem. Sens. Environ.* 202, 18–27.
- Gray, R., 1984. Vector quantization. *IEEE ASSP Mag.* 1, 4–29.
- Haut, J.M., Paoletti, M.E., Plaza, J., Plaza, A., 2018. Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines. *J. Real Time Imag. Process.* 15, 439–462.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780.
- Hoenser, T., Kuenzer, C., 2020. Object detection and image segmentation with deep learning on earth observation data: a review-part i: evolution and recent trends. *Rem. Sens.* 12, 1667.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets, 2017. Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv Preprint arXiv:1704.04861.
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS J. Photogrammetry Remote Sens.* 173, 24–49.
- Khanal, N., Matin, M.A., Uddin, K., Poortinga, A., Chishtie, F., Tenneson, K., Saah, D., 2020. A comparison of three temporal smoothing algorithms to improve land cover classification: a case study from Nepal. *Rem. Sens.* 12, 2888.
- Kingma, D.P., Ba, J., Adam, 2014. A Method for Stochastic Optimization, arXiv Preprint arXiv:1412.6980.
- Lemajic, S., Vajsova, B., Åstrand, P.J., 2018. New Sensors Benchmark Report on PlanetScope.
- 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. Photogrammetry Remote Sens.* 152, 166–177.
- Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2015. Deep learning earth observation classification using imagenet pretrained networks. *Geosci. Rem. Sens. Lett. IEEE* 13, 105–109.
- Noble, W.S., 2006. What is a support vector machine? *Nat. Biotechnol.* 24, 1565–1567.
- Oechaiyaphum, K., Ullah, H., Shrestha, R.P., Datta, A., 2020. Impact of long-term agricultural management practices on soil organic carbon and soil fertility of paddy fields in northeastern Thailand. *Geoderma Regional* 22, e00307.
- Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. *Rem. Sens. Environ.* 129, 122–131.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Rem. Sens. Environ.* 148, 42–57.
- Pan, S.J., Yang, Q., 2009. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359.
- Panboonyuen, T., Jitkajornwanich, K., Lawawirojwong, S., Srestasathien, P., Vateekul, P., 2019. Semantic segmentation on remotely sensed images using an enhanced global convolutional network with channel attention and domain specific transfer learning. *Rem. Sens.* 11, 83.
- Paoletti, M., Haut, J., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: a review. *ISPRS J. Photogrammetry Remote Sens.* 158, 279–317.
- Pashaei, M., Kamangir, H., Starek, M.J., Tissot, P., 2020. Review and evaluation of deep learning architectures for efficient land cover mapping with uas hyper-spatial imagery: a case study over a wetland. *Rem. Sens.* 12, 959.
- Pires de Lima, R., Marfurt, K., 2020. Convolutional neural network for remote-sensing scene classification: transfer learning analysis. *Rem. Sens.* 12, 86.
- Poortinga, A., Nguyen, Q., Tenneson, K., Troy, A., Bhandari, B., Ellenburg, W.L., Aekakkarunroj, A., Ha, L.T., Pham, H., Nguyen, G.V., et al., 2019a. Linking earth observations for assessing the food security situation in vietnam: a landscape approach. *Front. Environ. Sci.* 7, 186.
- Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., Saah, D., 2019b. Mapping plantations in Myanmar by fusing landsat-8, sentinel-2 and sentinel-1 data along with systematic error quantification. *Rem. Sens.* 11, 831.
- Poortinga, A., Aekakkarunroj, A., Kityuttachai, K., Nguyen, Q., Bhandari, B., Thwal, N.S., Priestley, H., Kim, J., Tenneson, K., Chishtie, F., et al., 2020. Predictive analytics for identifying land cover change hotspots in the mekong region. *Rem. Sens.* 12, 1472.
- Potapov, P., Tyukavina, A., Turubanova, S., Talero, Y., Hernandez-Serna, A., Hansen, M., Saah, D., Tenneson, K., Poortinga, A., Aekakkarunroj, A., et al., 2019. Annual continuous fields of woody vegetation structure in the lower mekong region from 2000-2017 landsat time-series. *Rem. Sens. Environ.* 232, 111278.
- Ronneberger, O., Fischer, P., Brox, T., U-net, 2015. Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.
- Saah, D., Johnson, G., Ashmall, B., Tondapu, G., Tenneson, K., Patterson, M., Poortinga, A., Markert, K., Quyen, N.H., San Aung, K., et al., 2019. Collect Earth: an Online Tool for Systematic Reference Data Collection in Land Cover and Use Applications. *Environmental Modelling & Software*.
- Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., San Aung, K., Markert, K.N., Clinton, N., Anderson, E.R., Cutter, P., et al., 2020. Primitives as building blocks for constructing land cover maps. *Int. J. Appl. Earth Obs. Geoinf.* 85, 101979.
- D. Saah, et al., **Land cover mapping in the 21st century: challenges and opportunities, update 1 (in review) 1.**
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4510–4520.
- Santiphop, T., Shrestha, R.P., Hazarika, M.K., 2012. An analysis of factors affecting agricultural land use patterns and livelihood strategies of farm households in kanchanaburi province, Thailand. *J. Land Use Sci.* 7, 331–348.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6, 60.
- Silalertruksa, T., Gheewala, S.H., 2018. Land-water-energy nexus of sugarcane production in Thailand. *J. Clean. Prod.* 182, 521–528.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv Preprint arXiv:1409, p. 1556.
- Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2019. Bigearthnet: a large-scale benchmark archive for remote sensing image understanding. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 5901–5904.
- Sun, J., Di, L., Sun, Z., Shen, Y., Lai, Z., 2019. County-level soybean yield prediction using deep cnn- lstm model. *Sensors* 19, 4363.
- Tenneson, K., Patterson, M., Jadin, J., Rosenstock, T., Mulia, R., Kim, J., Nguyen, Q., Poortinga, A., Phuong, N.M., Bogle, S., Dilger, J., Marlay, S., Tan, N.Q., Chishtie, F., Saah, D., 2021. Commodity-driven Forest Loss: A Study in Southeast Asia, p. 196.
- Tiwari, V., Matin, M.A., Qamer, F.M., Ellenburg, W.L., Bajracharya, B., Vadrevu, K., Rushi, B.R., Yusaf, W., 2020. Wheat area mapping in Afghanistan based on optical and sar time-series images in google earth engine cloud environment. *Front. Environ. Sci.* 8, 77.
- Trisurat, Y., Shirakawa, H., Johnston, J.M., 2019. Land-use/land-cover change from socio-economic drivers and their impact on biodiversity in nan province, Thailand. *Sustainability* 11, 649.
- Uddin, K., Shrestha, H.L., Murthy, M., Bajracharya, B., Shrestha, B., Gilani, H., Pradhan, S., Dangol, B., 2015. Development of 2010 national land cover database for the Nepal. *J. Environ. Manag.* 148, 82–90.
- Vadrevu, K.P., Lasko, K., Giglio, L., Justice, C., 2015. Vegetation fires, absorbing aerosols and smoke plume characteristics in diverse biomass burning regions of Asia. *Environ. Res. Lett.* 10, 105003. <https://doi.org/10.1088/1748-9326/10/10/105003>.
- Vermote, E., Tanré, D., Deuzé, J., Herman, M., Morcrette, J., Kotchenova, S., 2006. Second simulation of a satellite signal in the solar spectrum-vector (6sv), 6S User Guide Version 3, pp. 1–55.
- Walters, J.P., Archer, D.W., Sassenrath, G.F., Hendrickson, J.R., Hanson, J.D., Halloran, J.M., Vadas, P., Alarcon, V.J., 2016. Exploring agricultural production systems and their fundamental components with system dynamics modelling. *Ecol. Model.* 333, 51–65.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?. In: *Advances in Neural Information Processing Systems*, pp. 3320–3328.
- Young, N.E., Anderson, R.S., Chignell, S.M., Vorster, A.G., Lawrence, R., Evangelista, P.H., 2017. A survival guide to landsat preprocessing. *Ecology* 98, 920–932.
- Yu, J., Yin, T., Li, S., Hong, S., Peng, Y., 2019. Fast ship detection in optical remote sensing images based on sparse mobilenetv2 network. In: *International Conference on Genetic and Evolutionary Computing*. Springer, pp. 262–269.
- Yuttitham, M., Gheewala, S.H., Chidthaisong, A., 2011. Carbon footprint of sugar produced from sugarcane in eastern Thailand. *J. Clean. Prod.* 19, 2119–2127.
- Zhang, X., Han, L., Han, L., Zhu, L., 2020. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Rem. Sens.* 12, 417.
- Zhao, S., Liu, X., Ding, C., Liu, S., Wu, C., Wu, L., 2020. Mapping rice paddies in complex landscapes with convolutional neural networks and phenological metrics. *GIScience Remote Sens.* 57, 37–48.