Just-in-Time Biomass Yield Estimation with Multi-modal Data and Variable Patch Training Size

Patricia O'byrne, Patrick Jackman, Damon Berry, Thomas Lee, Michael French, Robert J. Ross

Just-in-time Biomass Yield Estimation with Multi-Modal Data and Variable Patch Training Size

Patricia O'Byrne[1 [0000−0001−6488−0201]], Patrick Jackman[1 [0000−0001−9540−5948]], Damon Berry[1 [0000−0002−2290−7661]], Thomas Lee[1], Michael French[2], and Robert J. Ross[1 [0000−0001−7088−273X]]

[1] Technological University Dublin, D07 ADY7 Dublin, Ireland
[2] Tanco Autowrap, R21 E278, Carlow, Ireland

Abstract The just-in-time estimation of farmland traits such as biomass yield can aid considerably in the optimisation of agricultural processes. Data in domains such as precision farming is however notoriously expensive to collect and deep learning driven modelling approaches need to maximise performance but also acknowledge this reality. In this paper we present a study in which a platform was deployed to collect data from a heterogeneous collection of sensor types including visual, NIR, and LiDAR sources to estimate key pastureland traits. In addition to introducing the study itself we address two key research questions. The first of these was the trade off of multi-modal modelling against a more basic image driven methodology, while the second was the investigation of patch size variability in the image processing backbone. This second question relates to the fact that individual images of vegetation and in particular grassland are texturally rich, but can be uniform, enabling subdivision into patches. However, there may be a trade-off between patch-size and number of patches generated. Our modelling used a number of CNN architectural variations built on top of Inception Resnet V2, MobileNet, and shallower custom networks. Using minimum Mean Absolute Percentage Error (MAPE) on the validation set as our metric, we demonstrate strongest performance of 28.23% MAPE on a hybrid model. A deeper dive into our analysis demonstrated that working with fewer but larger patches of data performs as well or better for true deep models – hence requiring the consumption of less resources during training.

Keywords: Deep Learning· CNN · Multimodal Processing · LiDAR · Patch-size· biomass · Precision Agriculture

1 Introduction

Optimised silage production is a key enabler in allowing milk and beef production levels to continue to grow in the face of the joint challenges of population growth [29] and diminishing resources [19]. While feeding trends vary from nation to nation, grass silage should ideally provide 20% to 30% of an animal's feed in a grass-fed environment[3]. When harvesting for silage, farmers benefit from knowing many different traits in the

[3] https://www.teagasc.ie/media/website/publications/2016/Teagasc-Quality-Grass-Silage-Guide.pdf

grassland being harvested. One of these is the biomass yield of their pasture [26]. Any systematic approach that can provide such information accurately and in a just-in-time manner can provide benefit to the silage optimisation process.

Beyond grassland management, food production and agriculture have embraced a wide variety of technological advances in recent years. Both proximal and remote sensing methods are now well recognised as important enablers for precision agriculture [10]. We are now at the point of developing sensor networks that can be applied to day-to-day farming activities. However, leveraging these technologies is not always straightforward, due to the costs involved in collecting datasets for training and the high levels of variability in environmental conditions. Even if we consider the simple application of so-called vegetation indices to biomass [30], we see that correlations between electromagnetic signatures and biomass vary according to plant phenology, ecology and sensing environment [23], rendering simple indices unreliable under diverse real world conditions.

In recent years there has been a trend towards methods that take a wider range of signals into account to overcome the limitations of vegetation indices. Image-based systems for example typically apply deep learning methods from the field of artificial intelligence to take advantage, not only of electromagnetic signatures, but also the true appearance of vegetation [6]. The challenge with Deep Learning driven methods tend to centre on the need to source large volumes of data to minimise bias in modelling. The application of transfer learning [16] to bootstrap model construction can be applied to help reduce training data requirements, but as always, a suitable source dataset is needed and when working beyond raw visible images, suitable source data sets are not always available [9].

Given the technical requirements of developing just-in-time pastureland assessment systems, in this paper we present a new study that has aimed to provide just-in-time robust estimation of biomass for silage production while considering a number of questions related to model optimisation under limited data constraint situations. In particular we explore a number of transfer learning and custom model designs that integrate data from a number of sensor types to estimate biomass yield. We specifically consider the relative impact of adding both LiDAR and NIR (near infrared) data to baseline image data, given that transfer learning from ImageNet driven models is straightforwardly available for RGB image data, but not for multispectral data. Alongside this, as grassland data is heterogeneous by nature, samples can be subdivided into patches to augment the dataset. We consider the impact that subdividing into different patch sizes has on model performance.

In light of more recent worries about the environmental impact of deep learning methods, we also ask what, if anything, we can learn regarding the costs of training and systems optimisation.

2 Related Work

We proceed by first setting out some key background concepts relevant to our study; including the sensor-based estimation of biomass, the application of machine learning

based methods to vegetation trait estimation, and issues related to the application of transfer learning in CNNs in overcoming limited training data constraints.

2.1 Remote Sensing of Vegetation

Traditionally, remote methods for estimating vegetation biomass have taken advantage of the relative absorption of different wavelengths of light through simple functions called vegetation indices (VIs). For example, for many decades the Normalized Difference Vegetation Index (NDVI) [18] has been applied to true remote sensing such as satellite data [28], but also hand-held sensing devices [2] and farm machinery mounted equipment [22]. Unfortunately, estimations from VIs such as NDVI are known to be highly subject to a range of conditions. NDVI correlates well to biomass early in the growing season, but saturates as vegetation becomes denser [27] and is influenced by soil exposure, topography, senescent vegetation and atmospheric contaminants [5]. Various modifications of basic vegetation indices have been proposed to overcome limitations; for example, canopy height has been used in conjunction with NDVI to extend its usefulness in dense vegetation [22].

Rather than focusing only on electromagnetic reflectance, more recent state-of-the-art remote sensing systems take advantage of the full visual analysis of vegetation. Such analysis is based on the principle that the texture and shape of a pasture canopy can help identify target characteristics [1]. Images may show droplets of water on the leaves, or evidence of drought, seed heads or leaf size, while canopy height is a strong indicator of biomass, as well as other vegetation traits. One such automated system that was developed to estimate vegetation cover and type, applied Local Binary Patterns (LBPs) to 1m x 1m photos of vegetation [15]. Deep learning has also been used to estimate NDVI from Sentinel satellites, even on a cloudy day [21], while the performance of a deep convolutional neural network (DCNN) was compared to that of conventional models built on feature extraction such as image segmentation, colour comparison of R/G, B/G and 2G-R-B without segmenting images [14]. This reflects a more general trend of applying state-of-the-art data processing methods, such as deep learning, in agriculture [11].

2.2 Deep Learning Architectures for Remote Sensing

In applying deep learning to image-centric precision agriculture data, there are a number of considerations to take into account in architectural choice. Although we might assume the application of CNNs as a backbone for image data channels, the range of CNN architectural variations is vast, as is the range of methods that can be applied to try to mitigate low volumes of training data. Two notable architectures that we build upon here are the Inception Networks [25,24] and MobileNet [7]. The Inception Networks are some of the most widely applied in computer vision at this time. The inception networks aim to overcome issues in scale invariance, through the application of heterogeneous kernel architectures and factorisation of large networks to produce predictors that give accurate estimations, while also incorporating skip-connections to increase network depth.

One challenge with the most sophisticated deep learning based image processing models such as Inception ResNet is that they are typically very large and subsequently require significant computational resources. This can be a challenge for deployment in fields such as agricultural machinery, where we may wish to limit the applied computational resources. Given such challenges, MobileNet, as its name would suggest, was designed for use on mobile devices, specifically for embedded computer vision applications using RGB data [7]. It uses a combination of multiple depth-wise and point-wise convolution layers to replace fewer, more resource-hungry convolutional layers. MobileNet V2 introduced residual connections to reinforce feature maps, and bottleneck layers to compress the data [20].

For problems targeting vegetation, where the cost of image labelling is high, in principle we can apply transfer learning to benefit from training on more generic image training datasets. ImageNet [3] has provided fertile ground for the production of pre-trained architectures for many years. Early examples in the agriculture domain include the use of AlexNet [12] and GoogLeNet [25] to detect plant diseases from a repository of plant health images [8,16]. More recently, another interesting work has shown how a more accurate classification of hyperspectral images for vegetation analysis has been developed using ResNet and transfer learning [9].

As indicated, transfer learning is frequently applied successfully in situations where there is a lack of training samples [4]. One caveat however on the use of transfer learning is that the data needs to be somewhat similar to the data on which the model has been trained; generally, this is straightforward for visual images, though differences of scale between images focused on vegetation and more generic training images such as people, places, and everyday objects can potentially cause challenges. Moreover, for many vegetation analysis tasks, such as our own, where multi-spectral data is to be used, there is a lack of pre-trained weights across different spectral bands. The result of this is that transfer learning cannot be applied straightforwardly in such cases without information loss.

3 Data Collection

Our analysis, presented later, has been constructed around a study we conducted to estimate grassland traits for a just-in-time estimation task. This study required a series of field and lab measurements, in order to build a dataset for pastureland traits and sensor data prior to silage production. This study built on an earlier pilot study presented previously [17], but was considerably more varied in terms of source data collections and the range and resolution of sensors deployed.

Data was collected using proximal sensors on a bespoke data collection trolley. Sensors included a four-channel JAI AD-130 GE camera to take RGB and NIR images, and a LiDAR-Lite v3HP device to record canopy height. Full details of the procedure used to perform data collection were presented in our previous study – including the procedure for data collection and labelling. In short however, a 50 cm^2 area was harvested and later weighted in lab conditions to determine biomass yield (Kg/Ha). Over 300 samples were collected from multiple locations around Ireland over 25 collection events during the growing season of 2019.

To prepare for modelling, a series of pre-processing steps were applied to our data to account for the relatively low data volumes for deep learning. The method we apply is similar to that used by Kubach et al. [13] in the field of histopathology. Firstly, a variant of 5-fold cross validation was applied. Specifically, the set of images was subdivided into five sets, labelled 1 to 5. These sets were recombined into five 4 + 1 datasets, in each case holding out one of the sets for validation and using the remaining four for training. Each sample had been recorded with image size of 964 x 1296 pixels for the RGB image and 966 x 1296 pixels for the NIR image. As the images are of grass, they are reasonably uniform in presentation. Because of this, we opted to subdivide the image data to augment the dataset, copying height and biomass values from the full image. Two approaches were taken. The first approach split each sample image into 156 x 156 pixel patches, giving a dataset of almost 10,000 patches. We named this dataset Small Patch, or SP. The second approach split the samples into 240 x 240 pixels, giving around 5,000 patches. This dataset is named Large Patch or LP. There are a lot more patches in SP than there are in LP, but there is not as much information in a given patch.

Every patch had an RGB image, an NIR image, a height scalar (cm) and a biomass scalar (Kg/Ha). As we see in the modelling later, whilst biomass was always the target, and RGB data was always used, the NIR data and height data were used in specific model variants to explore the relative advantage provided by these data channels to augment the central image information.

4 Modelling

In order to systematically investigate the relative merits of different architectural variations under our limited data constraint, we investigated a number of different models that varied in terms of backbone architecture, use of pre-training, data source modalities and patch sizes. In the following we summarise our model choices and their rationale.

4.1 Image Processing Backbone

Each of our models assume an image processing backbone for RGB image data. For this image processing backbone we have applied three variants which also differed in terms of whether transfer learning was applied. Our first variant is a shallow custom CNN (*Shallow*) which feeds the RGB image data into two 2D consecutive convolution layers using a 3x3 kernel, generating 24 feature maps, each with a relu activation function. 2x2 max pooling is then applied. This pattern is repeated, this time using 48 feature maps for both convolutional layers, before a final pooling layer. The output is flattened and fed into fully connected layers before a final target layer.

For the image processing backbone we also made use of both the Inception Renset V2 and MobileNet architectures. Inception Resnet V2 is well regarded for high performance. One factor of note here is the Inception Network's ability to mitigate scale variance in input images. However Inception Networks are expensive to train and smaller networks can be useful in domains such as our own. Therefore we also make use of the MobileNet architecture as a third candidate. Our overall network architecture based on both these backbones is similar to that used in the shallow network. An input

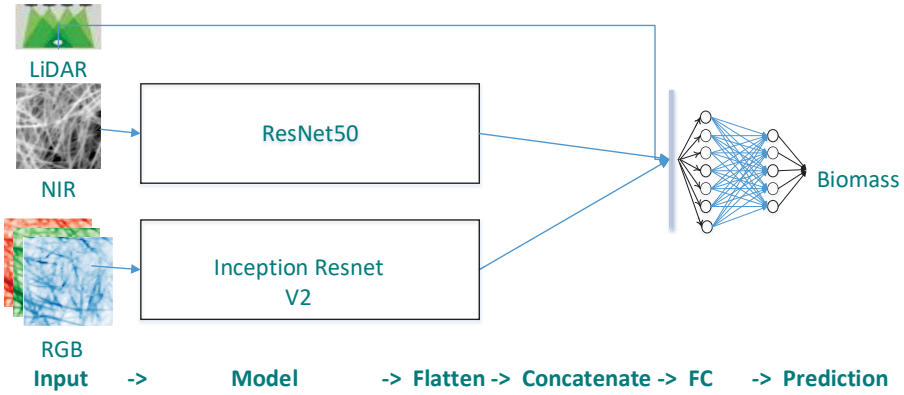

Figure 1: Hybrid NIR I L model. RGB and NIR image data are fed through CNN variants with outputs concatenated along with the scalar LiDAR information. The concatenated output is then fed through fully connected layers.

is passed through the backbone before being flattened and fed through fully connected layers and finally to an output layer.

Inception ResNet and MobileNet can of course be trained from scratch, but can also be used within a transfer learning methodology where pre-trained weights are loaded and then fine-tuned within the target application. For our image processing backbone we applied both transfer learning and 'from scratch' training methodologies. Models using ImageNet pre-trained weights are indexed with an 'I' for ImageNet. To facilitate the use of pre-trained weights, our input image, specifically input patches, were re-scaled to the required input image dimensions for both Inception Resnet and MobileNet, namely 299 x 299 and 224 x 224 pixels respectively.

4.2 Multi-Spectral and Multi-Sensor Analysis

Our second batches of analyses use a similar structure to the processing of visual data, but also introduce elements to incorporate NIR information into the processing backbone. For the shallow model we directly add a fourth channel to the visual input pipeline and 32 and 64 resultant image maps; this model is known as *Shallow NIR*. This four channel input was also used in one of the Inception ResNet models that trained from scratch *IncResNet NIR*. As ImageNet weights are trained for RGB data, we devised a further model design, using IncResNet with ImageNet weights on the RGB data and a parallel CNN architecture which focused on processing the NIR image before concatenating the results of that analysis to the output of the backbone image architecture. The NIR component was based on a ResNet-50 architecture. We refer to this integrated model as a 'Hybrid' model since it is neither strictly ResNet nor Inception ResNet.

Finally, we also introduce model variants that made use of LiDAR scalar data. Within an architecture the LiDAR distance estimates to the canopy were concatenated with the outputs of flattened CNN outputs before the concatenated vector was

fed through fully connected layers. Models which made use of the LiDAR data have the suffix 'L' affixed. To illustrate this approach, Figure 1 depicts the *Hybrid NIR I L* model. Here RGB data is fed into the pre-trained Inception ResNet V2 and NIR data is fed into ResNet50. The output of both is flattened and concatenated with LiDAR data, before going through the fully connected layers. Both Inception ResNet and ResNet50 use pre-trained ImageNet weights.

4.3 Influence of Patch Size

As described earlier, our data was pre-processed into two different size variants. One set, small patches (SP) was a dataset of ˜10,000 patches; each of size 156 x 156 pixels. The other set, large patches (LP), was a dataset of of ˜5,000 patches each of 240 x 240 pixels. Each of our models above were trained and tested against both the SP and LP data.

4.4 Training

In summary, ten models, cross-validated using 5-fold validation were trained and tested for each dataset. Two models used a shallow CNN, one with and one without NIR data. Four included Inception Resnet, three included MobileNet and the final one is a hybrid model that included both ResNet 50 and Inception ResNet. A full list is provided in Table 1. Where NIR data was used, the suffix NIR is shown. Where height data was used, the suffix L (for LiDAR) is shown. Where ImageNet weights were used to pre-train the model, the suffix I is shown. All models were trained with a mean square

Table 1: Model Architecture

Model	Includes	NIR	LiDAR	Weights
Shallow				
IncResNet	Inception ResNet			
IncResNet I	Inception ResNet			Y
MobileNet	MobileNet			
MobileNet I	MobileNet			Y
Shallow NIR		Y		
IncResNet NIR	Inception ResNet	Y		
IncResNet I L	Inception ResNet		Y	Y
MobileNet I L	MobileNet		Y	Y
Hybrid NIR I L	Inception ResNet and ResNet50	Y	Y	Y

error loss function and the Adam optimizer. Our metric for analysis is the Mean Absolute Percentage Error (MAPE). All models were trained for 300 epochs; early stopping was not applied in this case. All models were implemented using the Keras wrapper to Tensorflow. Training runs were carried out across two hardware platforms, one with a

Table 2: Minimum Mean Absolute Percentage Error values on Validation Data for each model variant for both Large and Small patch datasets. 'NIR' : NIR channel, 'I' : ImageNet pretraining, 'L': LiDAR scalars.

Model	Small	Large
Shallow	60.70%	62.84%
IncResNet	39.57%	39.42%
IncResNet I	37.55%	35.84%
MobileNet	41.26%	44.64%
MobileNet I	36.71%	33.31%
Shallow NIR	51.84%	57.02%
IncResNet NIR	41.44%	48.30%
IncResNet I L	29.60%	28.56%
MobileNet I L	30.93%	30.24%
Hybrid NIR I L	28.23%	31.11%

single Nvidia K40 GPU (A PowerEdge R730 with two 4-core Intel Xeon Processors @ 2.8 GHz, 512 GB RAM) and a higher-powered machine (Dell Dimension T5810 Tower with two 4-core Intel Xeon Processor @ 2.8 GHz, 4 GB RAM, a single RTX 2080 GPU).

5 Results and Discussion

Table 2 presents minimum Mean Absolute Percentage Error values for each model on a given patch size averaged over five runs. Referring to these results, the improvement between the shallow CNNs and the deeper Inception and MobileNet based networks is very evident. This in itself should not be surprising, but it does emphasise the need for relatively deep network architectures in IOT domains such as precision farming, despite the relative cost of needing more sophisticated hardware devices. In terms of the use of the deeper baseline models, the use of pre-trained ImageNet weights (i.e. those with a suffix 'I') is notable.

With respect to the incorporation of NIR and LiDAR data, the results are more mixed. The use of NIR data on the Shallow CNN improved performance in both patch size cases, but it proved to be an inhibitor when used with IncResNet. The additional use of NIR data with the IncResNet model did not improve performance and in fact, this poor performance was the impetus for developing the hybrid model, as with the hybrid model, both NIR and RGB images could be supplemented by pre-trained weights. Anywhere that LiDAR data was used, performance improved substantially.

As mentioned, in addition to enhancing model performance, we also wished to minimize resources. Therefore we also need to look at trade-offs between model performance and efficiency. As can be seen, the best performing model, with a minimum MAPE of 28.23% was the *Hybrid NIR I L* on the small patch data. This is not surprising, as this model was the largest and most resource intensive. It made full use of all of the data provided. However, the five folds ran for an average of almost 23 hours on our

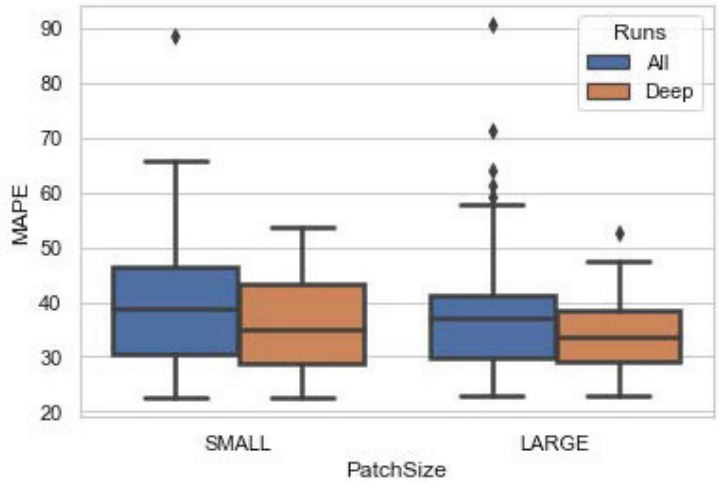

Figure 2: Distribution of minimum MAPE over all runs and deep runs

higher spec machine. The next best performer was *IncResNet I L* on the large patch data with an average minimum MAPE of 28.58%. The average runtime over the five runs of this model, on the same machine, was 8 hours. Whilst the difference in performance is very small, it is worth noting that the deeper model using extra NIR data and with double the number of patches does not seem to perform significantly better than the more economical one. Another question we wish to ask is whether there was a significant difference in outcome between a large number of small patches or a smaller number of large patches given that such variation is possible with our dataset type. To test this we applied a t-test across all runs, split into SP and LP groups. In all, there were 50 runs, 5 for each model on the small patch (SP) data and 5 for each model on the large patch (LP) data. The t-test on minimum MAPE over all runs showed 0.78, with a p-value of 0.439, suggesting that there was in fact no difference between the use of the SP and LP datasets. The distribution of MAPE values is shown in Figure 2 as All Runs. Note that both small and large patch results have outliers at around 90%. However, the shallow CNN results gave averages between 51.84% and 62.84% and can be thought of as more greedy for samples as they have no prior knowledge to build upon. If we repeat our analysis with *Shallow* and *Shallow NIR* models omitted, the results change considerably. The distribution of values for the deep runs are also shown in Figure 2, this time as Deep Runs. On just the 40 deeper runs, the mean is 35.66 for SP and 33.71 for LP, with a t-test of 2.55 and p-value=0.016.

Our interpretation overall is that choice of patch size matters, but in complex ways. For networks trained from scratch there is advantage in having a large number of patches available to progress training quickly, given that there is no pre-trained knowledge to build upon. However, when using pre-trained networks, this advantage disappears. In fact in many cases larger samples, though fewer in number, perform as well as,

if not better than a larger number of small samples. This suggests the pre-trained networks are better able to take advantage of the information present in the larger images. From an environmental perspective we point out that there is a training cycle advantage here to be noted. Training with fewer larger inputs will naturally result in fewer training batches and hence an energy / cost saving relative to training with a larger number of small batches. In both cases the potential for over-fitting the data remains, but here the focus on the validation rather than training metrics on a suitably split dataset help to minimise this potential.

6 Conclusion

In this paper we have presented an overview of a data collection platform for the assessment of biomass in a precision farming context. Our modelling approach has built on previous findings based on the use of vegetation indices and image processing. Whilst traditionally, NIR data has been considered an essential ingredient in predicting biomass, in our analyses with deeper models it has been eclipsed by the use of RGB data using pre-trained ImageNet weights, supplemented by LiDAR data. When augmenting the dataset by subsetting image samples, we found that running a model on large patch data using just RGB data and LiDAR took fewer resources with negligible performance loss.

Acknowledgements The authors wish to thank Enterprise Ireland and Tanco Autowrap Ltd. for the their support for this paper under Innovation Partnership Project IP2018-0728 which is co-funded by the European Regional Development Fund.

References

1. Abadi, M., Capelle-Laizé, A.S., Khoudeir, M., Combes, D., Carré, S.: Grassland Species Characterization for Plant Family Discrimination by Image Processing. In: Image and Signal Processing. pp. 173–181. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg (Jun 2010). https://doi.org/10.1007/978-3-642-13681-8_21, https://link.springer.com/chapter/10.1007/978-3-642-13681-8_21
2. Abičić, I., Lalić, A., Galić, V., Mlinarić, S., Begović, L.: Application of biomass sensor in the winter barley selection. Zbornik radova 54. Hrvatskog i 14. međunarodnog simpozija agronoma p. 179 (Feb 2019). https://doi.org/https://doi.org/10.1007/978-3-642-13681-8-21, https://www.bib.irb.hr/987036?rad=987036
3. Deng, J., Dong, W., Socher, R., Li, L.J., Kai Li, Li Fei-Fei: ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255. IEEE, Miami, FL (Jun 2009). https://doi.org/10.1109/CVPR.2009.5206848, https://ieeexplore.ieee.org/document/5206848/
4. Ferreira, C., Melo, T., Sousa, P., Meyer, M., Shakibapour, E., Costa, P., Campilho, A.: Classification of Breast Cancer Histology Images Through Transfer Learning Using a Pre-trained Inception Resnet V2. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **10882 LNCS**, 763–770 (2018). https://doi.org/10.1007/978-3-319-93000-8-86
5. Garroutte, E.L., Hansen, A.J., Lawrence, R.L.: Using NDVI and EVI to Map Spatiotemporal Variation in the Biomass and Quality of Forage for Migratory Elk in the Greater Yellowstone

Ecosystem. Remote Sensing **8**(5), 404 (May 2016). https://doi.org/10.3390/rs8050404, http://www.mdpi.com/2072-4292/8/5/404
6. Halstead, M., McCool, C., Denman, S., Perez, T., Fookes, C.: Fruit Quantity and Ripeness Estimation Using a Robotic Vision System. IEEE Robotics and Automation Letters **3**(4), 2995–3002 (Oct 2018). https://doi.org/10.1109/LRA.2018.2849514, conference Name: IEEE Robotics and Automation Letters
7. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861 [cs] (Apr 2017), http://arxiv.org/abs/1704.04861, arXiv: 1704.04861
8. Hughes, D.P., Salathe, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060 [cs] (Nov 2015), http://arxiv.org/abs/1511.08060, arXiv: 1511.08060
9. Jiang, Y., Li, Y., Zhang, H.: Hyperspectral Image Classification Based on 3-D Separable ResNet and Transfer Learning. IEEE Geoscience and Remote Sensing Letters pp. 1–5 (2019). https://doi.org/10.1109/LGRS.2019.2913011
10. Jiménez, J.d.l.C., Leiva, L., Cardoso, J.A., French, A.N., Thorp, K.R.: Proximal sensing of Urochloa grasses increases selection accuracy. Crop and Pasture Science **71**(4), 401–409 (May 2020). https://doi.org/10.1071/CP19324, https://www.publish.csiro.au/cp/CP19324, publisher: CSIRO PUBLISHING
11. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: A survey. Computers and Electronics in Agriculture **147**, 70–90 (Apr 2018). https://doi.org/10.1016/j.compag.2018.02.016, https://linkinghub.elsevier.com/retrieve/pii/S0168169917308803
12. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (May 2017). https://doi.org/10.1145/3065386, http://dl.acm.org/citation.cfm?doid=3098997.3065386
13. Kubach, J., Muhlebner-Fahrngruber, A., Soylemezoglu, F., Miyata, H., Niehusmann, P., Honavar, M., Rogerio, F., Kim, S.H., Aronica, E., Garbelli, R., Vilz, S., Popp, A., Walcher, S., Neuner, C., Scholz, M., Kuerten, S., Schropp, V., Roeder, S., Eichhorn, P., Eckstein, M., Brehmer, A., Kobow, K., Coras, R., Blumcke, I., Jabari, S.: Same same but different: A Web-based deep learning application revealed classifying features for the histopathologic distinction of cortical malformations. Epilepsia **61**(3), 421–432 (2020). https://doi.org/https://doi.org/10.1111/epi.16447, https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.16447, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.16447
14. Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A.: Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing **152**, 166–177 (Jun 2019). https://doi.org/10.1016/j.isprsjprs.2019.04.015, http://www.sciencedirect.com/science/article/pii/S0924271619301108
15. McCool, C., Beattie, J., Milford, M., Bakker, J.D., Moore, J.L., Firn, J.: Automating analysis of vegetation with computer vision: Cover estimates and classification. Ecology and Evolution **8**(12), 6005–6015 (2018). https://doi.org/10.1002/ece3.4135, https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.4135
16. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using Deep Learning for Image-Based Plant Disease Detection. Frontiers in Plant Science **7** (2016). https://doi.org/10.3389/fpls.2016.01419, https://www.frontiersin.org/articles/10.3389/fpls.2016.01419/full
17. O'Byrne, P., Jackman, P., Berry, D., Lee, T., French, M., Ross, R.J.: Transfer Learning Performance for Remote Pastureland Trait Estimation in Real-time Farm Monitoring. In: IGARSS IEEE International Geoscience and Remote Sensing Symposium. Brussels, Belgium (Jul 2021)

18. Rouse, Jr., J.W., Haas, R.H., Schell, J.A., Deering, D.W.: Monitoring Vegetation Systems in the Great Plains with ERTS. NASA Special Publication **351**, 309 (1974), https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022614.pdf
19. Sadhukhan, J., Dugmore, T.I.J., Matharu, A., Martinez-Hernandez, E., Aburto, J., Rahman, P.K.S.M., Lynch, J.: Perspectives on "Game Changer" Global Challenges for Sustainable 21st Century: Plant-Based Diet, Unavoidable Food Waste Biorefining, and Circular Economy. Sustainability **12**(5), 1976 (Jan 2020). https://doi.org/10.3390/su12051976, https://www.mdpi.com/2071-1050/12/5/1976, number: 5 Publisher: Multidisciplinary Digital Publishing Institute
20. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2 Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4510–4520. IEEE (2018)
21. Scarpa, G., Gargiulo, M., Mazza, A., Gaetano, R.: A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. Remote Sensing **10**(2), 236 (2018), /paper/A-CNN-Based-Fusion-Method-for-Feature-Extraction-Scarpa-Gargiulo/0ce69a9226f4d9eb8058065b95196f1265891a8f
22. Schaefer, M.T., Lamb, D.W.: A Combination of Plant NDVI and LiDAR Measurements Improve the Estimation of Pasture Biomass in Tall Fescue (Festuca arundinacea var. Fletcher). Remote Sensing **8**(2), 109 (Feb 2016). https://doi.org/10.3390/rs8020109, http://www.mdpi.com/2072-4292/8/2/109
23. Sims, D.A., Gamon, J.A.: Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sensing of Environment **81**(2-3), 337–354 (Aug 2002). https://doi.org/10.1016/S0034-4257(02)00010-X, http://linkinghub.elsevier.com/retrieve/pii/S003442570200010X
24. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)
25. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–9. IEEE, Boston, MA, USA (Jun 2015). https://doi.org/10.1109/CVPR.2015.7298594, http://ieeexplore.ieee.org/document/7298594/
26. Tan, A.E., Richards, S., Sarrabezolles, L., Platt, I., Woodhead, I.: Proximal soil moisture sensing of dairy pasture. In: 2014 IEEE Conference on Antenna Measurements Applications (CAMA). pp. 1–4 (Nov 2014). https://doi.org/10.1109/CAMA.2014.7003402
27. Tanaka, S., Kawamura, K., Maki, M., Muramoto, Y., Yoshida, K., Akiyama, T.: Spectral Index for Quantifying Leaf Area Index of Winter Wheat by Field Hyperspectral Measurements: A Case Study in Gifu Prefecture, Central Japan. Remote Sensing **7**(5), 5329–5346 (May 2015). https://doi.org/10.3390/rs70505329, https://www.mdpi.com/2072-4292/7/5/5329
28. Townshend, J.R.G., Justice, C.O.: Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. International Journal of Remote Sensing **9**(2), 187–236 (Feb 1988). https://doi.org/10.1080/01431168808954847, https://doi.org/10.1080/01431168808954847
29. UN: World Population Prospects The 2015 Revision Key Findings and advance tables. Tech. Rep. ESA/P/WP.241, United Nations, Department of Economic and Social Affairs, Population Division, New York City (2015), https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf
30. Xue, J., Su, B.: Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. Journal of Sensors (2017). https://doi.org/10.1155/2017/1353691, https://www.hindawi.com/journals/js/2017/1353691/