YIELD ESTIMATION of SUNFLOWER PLANT with CNN and ANN USING SENTINEL-2

O. G. Narin[1]*, A. Sekertekin[2], A. Saygin[3], F. Balik Sanli[4], M. Gullu[1]

[1]Department of Geomatics Engineering, Afyon Kocatepe University, Turkey – (gokberknarin@aku.edu.tr, mgullu@aku.edu.tr)
[2]Department of Geomatics Engineering, Cukurova University, Turkey - asekertekin@cu.edu.tr
[3]Department of Geomatics Engineering, Hacettepe University, Turkey - sayginabdikan@hacettepe.edu.tr
[4]Department of Geomatics Engineering, Yildiz Technical University, Turkey - fbalik@yildiz.edu.tr

KEY WORDS: Yield Estimation, Sentinel-2, CNN, ANN, NDVI, NDVIred.

ABSTRACT:

Due to food security and agricultural land management, it is crucial for decision makers and farmers to predict crop yields. In remote sensing based agricultural studies, spectral resolutions of satellite images, as well as temporal and spatial resolution, are important. In this study, we investigated whether there is a relationship between the Normalized Different Vegetation Index (NDVI) and Normalized Different Vegetation Index Red-edge (NDVIred) indices derived from the Sentinel-2 satellite. In addition, the efficiency of linear regression, Convolutional Neural Network (CNN), and Artificial Neural Network (ANN) techniques are examined with the use of indices in yield estimation. In this context, yield data of 48 sunflower parcels were obtained in 2018. The obtained results showed that both NDVI and NDVIred can be used to estimate the yield of sunflowers. The best results were obtained from the combination of the NDVI and the CNN technique with the RMSE equal to 20,874 Kg/da on 30 June 2018. Concerning the results, although there is not much superiority between the two indices, the best results were generally obtained from CNN as the method.

1. INTRODUCTION

Agriculture has a key role in economic development of the countries, increasing food security and social well-being, and limiting the impact of the climate change (Mok et al., 2014; Byerlee et al., 2009; Palatnik and Roson, 2012). The food scenario of the world has been changing so fast due to the increase in global population. Since the arable land resources are limited and the number of these resources has been decreasing day by day, the pressure on presently productive land is greater than ever before (Seelan et al., 2003). It is also projected that the arable land will decline to about 0.15 ha per capita by 2050 (Lal, 1991). Moreover, the global demand for food is projected to increase by 1.5–2 times between 1990 and 2030 (Daily et al., 1998). Therefore, it is crucial to increase the productivity and the number of the agricultural lands as well as to develop a sustainable development strategy for food security.

Food security is one of the critical issues for every government and country. In this context, accurate and timely crop yield estimation is extremely valuable for decision makers and crop producers. Various methods are available for crop yield estimation at different scales, from plot to continental scale. Direct methods, which refer to the ground measurements, represent reliable yield estimations; however, they are not cost- and time-effective, and it is too difficult to apply these methods over large areas (Burke and Lobell, 2017). Crop growth models, which include ecophysiological processes to simulate crop growth, development, and yields according to soil characteristics, agricultural practices, and meteorological data, are the other ways of crop yield estimation (Leroux et al., 2019). On the other hand, remote sensing (RS) technology can provide higher field coverage and functional preliminary information about the growing crops for a better productivity and yield estimation (Narin and Abdikan, 2020).

Nowadays, RS data, acquired from ground-based, airborne-based, and/or space-based platforms, have been widely used in agricultural applications such as weed mapping (Lamb and Brown, 2001), water content mapping of crops (Tilling et al., 2006), soil properties mapping (Barnes et al., 2003), and yield estimation (Fieuzal et al., 2017; Ali et al., 2019; Kayad et al., 2019; Narin and Abdikan, 2020; Ouattara et al., 2020). Concerning the yield estimation studies with RS data, various models such as statistical, numeric, and/or machine learning (ML) approaches are utilized to define the relationship between yield and other variables that are related to the yield. Among these methods, ML techniques generally provide better performance and higher accuracy compared to the conventional methods since they learn to model complexity through training (Kayad et al., 2019). More details about machine learning approaches for crop yield prediction can be found in the review paper of Chlingaryan et al. (2018).

This study has three main objectives; (1) to estimate sunflower yield using Sentinel-2 based Vegetation Indices (VIs) and three methods including two machine learning approaches, namely, Artificial Neural Network (ANN) and Convolutional Neural Network (CNN), and linear regression. (2) To investigate the performance of three methods in yield estimation with satellite imagery. (3) To examine the effectiveness of two vegetation indices, namely, Normalized Difference Vegetation Index (NDVI) and NDVI red-edge (NDVIred), in yield estimation with the corresponding methods. As a crop type, sunflower, which is one of the four most important annual crops in the world grown for edible oil (Putt, 1977), was considered in this study. In addition, the study was conducted in an agricultural area in Zile district,Tokat province, Turkey, at plot scale.

2. MATERIALS AND METHODS

In this study, the relationship between yields of the sunflower plant and Sentinel-2 based vegetation indices (NDVI and NDVIred) was investigated. A field survey was conducted for the determination of parcel boundaries, and yield information was obtained from the farmers. Sentinel-2 images of 30 June, 8 July, and 10 July were selected in the study. The reason for choosing these images is that the highest correlation was seen in these three

https://doi.org/10.5194/isprs-archives-XLVI-4-W5-2021-385-2021 |

dates in our previous study (Narin and Abdikan, 2020). To identify the relationship between yield and vegetation indices at the plot scale, the pixels close to the parcel boundaries were not chosen, and the reflection from the adjacent parcels was tried to be minimized. Between the selected pixels and the yield, linear regression, CNN, and ANN methods are used to estimate the seasonal yield. The cross-validation (4-fold) technique was used for data of 48 yield estimation (36 fields for training and 12 fields for testing). In cross-validation, training and testing groups are changed four times and their mean is considered for evaluation. The each approach was evaluated using RMSE and R^2 calculated from the equation 1 and equation 2 given below. Besides, the workflow of the methodology is given in figure 1.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{act} - y_i^{cal})} \tag{1}$$

where n is the number of parcels, y_i^{act} is the known value, and y_i^{cal} is the value produced by the linear regression, CNN, and ANN.

$$R^2 = \frac{\sum(y_i^{act}-\overline{y_i^{act}})(y_i^{cal}-\overline{y_i^{cal}})}{\sqrt{\sum(y_i^{act}-\overline{y_i^{act}})^2(y_i^{cal}-\overline{y_i^{cal}})^2}} \tag{2}$$

where y_i^{act} is the measured sunflower yield and y_i^{cal} is estimated sunflower yield, $\overline{y_i^{act}}$ and $\overline{y_i^{cal}}$ the average measured and estimated sunflower yield.

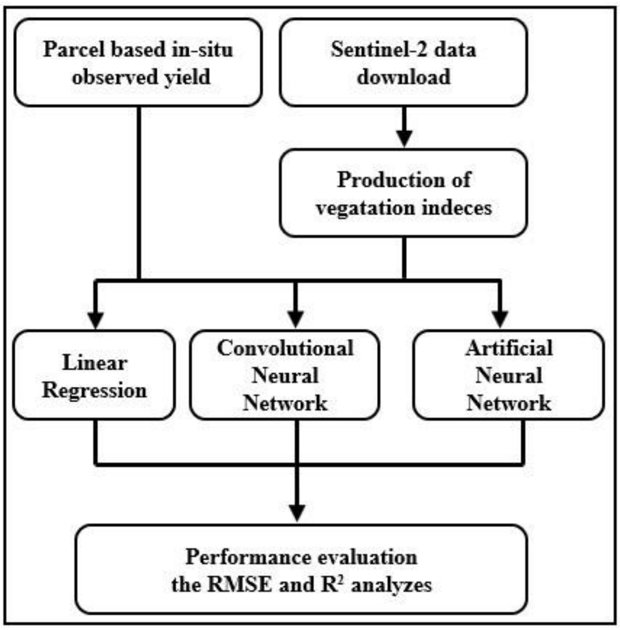

Figure 1. The workflow of the methodology.

2.1 Study Area and Sunflower Plants

The study area is located in Zile district of Tokat province, Turkey (Figure 2). Zile district is lying on 70 km west of the center of Tokat, between 35° 25' and 36° 6' east longitudes, 40° 4' and 40° 26' north latitudes, and its average altitude is 710m above sea level. It has a continental climate. The mean annual temperature is 11.7 °C and the mean annual rainfall is 436 mm in the Zile district (Climate Data). The soil structure in the study area is clayey and the clay contents of the parcels vary between 45% and 55%, and the amount of organic matter in the soil varies between 7.13% and 8.23%. Sunflower plants are planted in the region in early May and harvested after about 4 months. 48 sunflower parcels were observed and the average yield value for sunflower is about 321 kg/ha in the study area. The yield information in the study was obtained from the farmers.

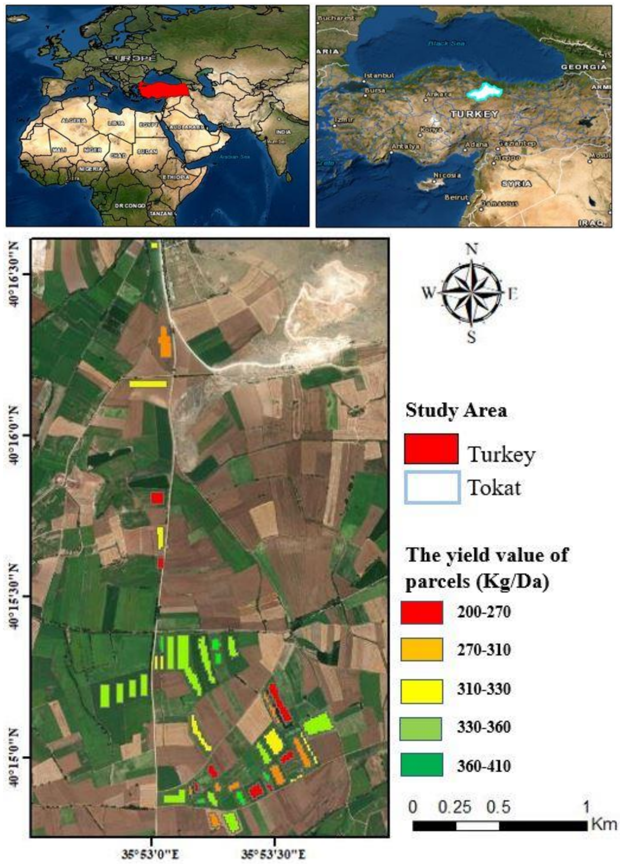

Figure 2. Location map of the study area and test sites.

2.2 Sentinel-2 Satellite Data and Vegetation Indices

In our study, we used the Sentinel-2 mission that allowed optical observations of the world since March 7, 2017. The mission is now a constellation with two satellites, Sentinel-2A and Sentinel-2B, and these satellites have 13 spectral bands with different spatial resolutions. To generate the indices, we used band 4 (Red) and band 8 (NIR) with a spatial resolution of 10 m and band 5 (Red-edge) with a spatial resolution of 20 m (Table 1). We used the Bottom of Atmosphere (BOA) corrected Level-2A images (ESA). In addition, Sentinel Application Platform (SNAP) software was used for VIs calculation and pixel resampling processes.

Equation	VIs Name	References
(B8-B4)/(B8+B4)	NDVI	(Rouse et al. 1974)
(B8-B5)/(B8+B5)	NDVIred	(Gitelson and Merzlyak 1994)

Table 1. VIs extraction from Sentinel-2 satellite data.

2.3 Linear Correlation

In order to estimate the yield, firstly, linear regression functions were created between the VIs and yield values. Since the correlation coefficient represents the strength and relationship between two variables, the linear regression function uses Equation 3 to express this relationship;

$$y = a + bx + \varepsilon \tag{3}$$

where y is the dependent and x is the independent variable, a is the slope, b is the intercept and ε is the error (Yan and Su 2009).

2.4 Artificial Neural Network

Secondly, for the yield estimation, ANN approach was applied. ANNs are simulated based on biological neural networks. They consist of multi interconnected neurons with coefficients to compose a neural structure (Inyurt and Sekertekin 2019).

In this study, a backpropagation approach is used to train the multi-layer perceptron feed-forward network. While training the function, the most appropriate network structure was tried to be determined by the trial-and-error method in order to improve the transitions from the hidden layer to the output layer. The network structure is [1,18,1].

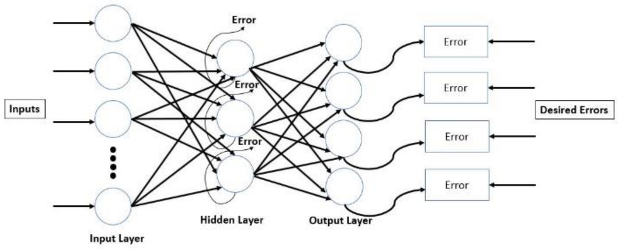

Figure 3. BPANN structure (Gullu and Narin 2019)

2.5 Convolutional Neural Network

In the last step, a deep learning algorithm Convolutional Neural Network (CNN) approach was tested. It was used to forecast different variables such as yield prediction of winter wheat (Mu et al. 2019), corn and soybean (Khaki et al. 2020), wheat and barley (Nevavuori et al. 2019), rice (Yang 2019), and also for hazard analysis such as forest fire modeling (Zhang et al. 2019), and landslide susceptibility analysis (Sameen et al. 2020). The CNN application was executed in R software (CNN Code). In our study, the Keras sequential model was used with a one-dimensional CNN model. Relu function is used for activation. Concerning the model variables, filters and units were 256 and 1024, respectively, and Adam optimization algorithm is used to optimize in CNN model. The number of the epochs was chosen as 250 to train the data of the model.

3. RESULTS AND DISCUSSION

The results showed that the CNN approach provided the best estimation compared to ANN and linear regression methods (Table 2). Among the three acquisitions, the lowest RMSE value was determined with the image acquired on 30th June. Once compared the indices for 30th June results, NDVI provided a slightly better result than NDVIred for CNN. In addition, NDVI provided better results with linear regression and ANN results compared to the NDVIred data. The images acquired in July provided similar results that the RMSE ranges between 27 kg/da and 31 kg/da. Between two dates image acquired on 10th July estimated better yield values than 8th July. In the case of July, NDVIred resulted a lower RMSE values against NDVI. Except for the results of 8th July, CNN gave better results than the other approaches. The results of ANN provided the second best estimation after the CNN method.

Coefficient of determination (R^2) results also showed that the CNN approach has higher values compared to ANN and linear regression (Table 3). NDVIred has higher R^2 value than NDVI on all acquisition dates, but they are almost the same in June. Among the three dates, 30th June has higher R^2 values for all methods and indices.

The results of the study determined a similar trend compared to the previous study Narin and Abdikan (2020). Additionally, it is also noticed that the inflorescence emergence stage of the sunflower is a better period to estimate the yield of sunflower for the region. However, in this study, ANN and CNN estimated higher R^2 and lower RMSE values than linear regression. It is also indicated that the NDVIred can be an alternative index to NDVI.

Date		NDVI			NDVIred		
		RMSE (Kg/Da)			RMSE (Kg/Da)		
		Linear Regression	CNN	ANN	Linear Regression	CNN	ANN
30 June	Grup 1	15.995	13.247	15.898	20.676	17.530	19.105
	Grup 2	21.989	20.603	18.828	21.667	18.264	21.079
	Grup 3	19.223	17.227	17.584	20.082	19.041	20.549
	Grup 4	33.34	32.418	31.423	32.265	29.655	29.821
	Mean	22.637	**20.874**	20.933	23.673	**21.123**	22.638
8 July	Grup 1	20.788	20.738	20.119	19.569	19.180	19.203
	Grup 2	29.938	28.209	22.932	27.610	26.046	25,771
	Grup 3	33.728	32.587	35.711	33.298	32.004	35.495
	Grup 4	44.822	42.687	44.116	41.282	40.601	40.326
	Mean	32.319	31.055	**30.719**	30.440	**29.458**	30,199
10 July	Grup 1	18.377	17.955	18.357	19.158	17.360	17.358
	Grup 2	26.275	24.846	25.814	23.183	21.577	22.055
	Grup 3	36.928	33.326	39.600	35.512	32.397	34.438
	Grup 4	42.939	41.761	41.414	39.473	36.986	38.853
	.Mean	31.130	**29.472**	31.139	29.332	**27.080**	28.176

Table 2. RMSE results obtained from the methods and indices.

Date	R^2					
	NDVI			NDVIred		
	Linear Regression	CNN	ANN	Linear Regression	CNN	ANN
30 June	0.748	0.791	0.785	0.735	**0.792**	0.763
8 July	0.495	0.535	0.552	0.553	0.592	0,567
10 July	0.520	0.582	0.528	0.578	0.654	0.610

Table 3. R^2 results obtained from the methods and indices.

4. CONCLUSIONS

Accurate determination of agricultural parameters is important for monitoring policies and rapid decision-making on crops. The yield of crops for each parcel is one of the critical parameters and it conventionally can be determined after the harvest of the crop. However, its earlier estimation is important to manage water usage and to take precautions when necessary during their growing periods. Thus, the use of remote sensing data enable monitoring agricultural fields for long periods, and enables estimating the yield information before the harvesting period of the crops.

In this study, the contribution of linear regression, ANN, and CNN were tested to estimate crop yield with satellite image based vegetation indices. For the analysis, the sunflowers are monitored during the growing period, and three images were considered for the estimation according to previous studies. Although the dataset was limited, we were able to train the network in both CNN and ANN. It is useful to try the method in places where there are more data sets. It is concluded that the CNN gave slightly better results compared to both the ANN and the linear regression. Both NDVI and NDVIred derived from Sentinel-2 data provided good results for all combinations. The best estimation is determined for 30th of June image that is corresponding to the inflorescence emergence stage of the sunflower. The study showed that the Sentinel-2 based NDVI and NDVIred can be used to determine the yield of the sunflower before the harvesting period.

ACKNOWLEDGEMENTS

The authors would like to thank ESA (European Space Agency) for providing free access to the Sentinel-2 satellite data and SNAP tool used in the present study. The authors would like to thank research assistant MSc. Ahmet DELEN, and the owners of the sunflower fields.

REFERENCES

Ali, A., Martelli, R., Lupia, F., & Barbanti, L. (2019). Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. Remote Sensing, 11(20), 2384.

Barnes, E.M., Sudduth, K.A., Hummel, J.W., Lesch, S.M., Corwin, D.L., Yang, C., Daughtry, C.S.T., Bausch, W.C., 2003. Remote- and ground-based sensor techniques to map soil properties. Photogramm. Eng. Remote Sens. 6, 619–630.

Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proceedings of the National Academy of Sciences, 114(9), 2189-2194.

Byerlee, D., De Janvry, A., & Sadoulet, E. (2009). Agriculture for development: Toward a new paradigm. Annu. Rev. Resour. Econ., 1(1), 15-31.

Chan, K.L., Qin K., 2017: Biomass burning related pollution and their contributions to the local air quality in Hong Kong. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-2/W7, 29-36. doi.org/10.5194/isprs-archives-XLII-2-W7-29-2017.

Climate data - Climate data in cities worldwide. https://tr.climate-data.org/asya/tuerkiye/tokat/zile-15834 (July 2021)

CNN Code - How to Fit Regression Data with CNN Model in R https://www.datatechnotes.com/2020/01/how-to-fit-regression-data-with-cnn.html (July 2021)

Daily, G. C., Dasgupta, P., Bolin, B., Crosson, P., Guerny du, J., Ehrlich, P. R., Folke, C., Jansson, A. M., Jansson, B. -O., Kautsky, N., Kinzig, A., Levin, S., Ma¨ler, K. -G., Pinstrup-Andersen, P., Siniscalco, D., & Walker, B. (1998). Food production, population growth, and the environment. Science, 281, 1291– 1292.

Dubayah, R.O., Swatantran, A., Huang, W., Duncanson, L., Tang, H.,Johnson, K., Dunne, J.O., Hurtt, G.C., 2017. CMS: LiDAR-derived Biomass, Canopy Height and Cover, Sonoma County, California, 2013. ORNL DAAC, Oak Ridge, Tennessee, USA. doi.org/10.3334/ORNLDAAC/1523.

ESA – European Space Agency https://earth.esa.int/documents/247904/685211/Sentinel-2_User_Handbook (July 2021)

Förstner, W., Wrobel, B., 2016: Photogrammetric Computer Vision. Springer Nature, Cham.

Fieuzal, R., Sicre, C. M., & Baup, F. (2017). Estimation of sunflower yield using a simplified agrometeorological model controlled by optical and SAR satellite data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(12), 5412-5422.

Gago-Silva, A., 2016. GRASS GIS in Grid Environment. doi.org/10.6084/m9.figshare.3188950.

Gitelson, A. A., Merzlyak, M. (1994). Spectral reflectance changes associated withautumn senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral features and relation to chlorophyll estimation. J. Plant Physiol. 143(3), 286–292.

GRASS Development Team, 2015. Geographic Resources Analysis Support System (GRASS) Software, Version 6.4. Open Source Geospatial Foundation. grass.osgeo.org (1 June 2017).

GRASS Development Team, 2017. Geographic Resources Analysis Support System (GRASS) Software. Open Source Geospatial Foundation. grass.osgeo.org (20 September 2017).

Gullu, M., Narin, O., G. (2019). Georeferencing of the Nile River in Piri Reis 1521 map, Using artificial neural network method. Acta Geod Geophys, 54(3), 387-401 https://doi.org/10.1007/s40328-019-00255-7

Inyurt, S., Sekertekin, A. Modeling and predicting seasonal ionospheric variations in Turkey using artificial neural network (ANN). Astrophys Space Sci 364, 62 (2019). https://doi.org/10.1007/s10509-019-3545-9

Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. Remote Sensing, 11(23), 2873.

Khaki S, Wang L and Archontoulis SV (2020) A CNN-RNN Framework for Crop Yield Prediction. Front. Plant Sci. 10:1750. doi: 10.3389/fpls.2019.01750

Lal, R. (1991). Soil structure and sustainabilty. Journal of Sustainable Agriculture, 4, 67– 92.

Lamb, D.W., Brown, R.B., 2001. PA—Precision agriculture. J. Agric. Eng. Res. 78, 117–125.

Lennert, M., GRASS Development Team, 2017. Addon i.segment.stats. Geographic Resources Analysis Support System (GRASS) Software, Version 7.2, Open Source Geospatial Foundation. grass.osgeo.org/grass7/manuals/addons/i.segment.stats (1 June 2017).

Leroux, L., Castets, M., Baron, C., Escorihuela, M. J., Bégué, A., & Seen, D. L. (2019). Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. European Journal of Agronomy, 108, 11-26.

Maas, A., Rottensteiner, F., Heipke, C., 2017. Classification under label noise using outdated maps. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-1/W1, 215-222. doi.org/10.5194/isprs-annals-IV-1-W1-215-2017.

Michalis, P., Dowman, I., 2008: A Generic Model for Along-Track Stereo Sensors Using Rigorous Orbit Mechanics. Photogrammetric Engineering & Remote Sensing 74(3), 303-309.

Mok, H. F., Williamson, V. G., Grove, J. R., Burry, K., Barker, S. F., & Hamilton, A. J. (2014). Strawberry fields forever? Urban agriculture in developed countries: a review. Agronomy for sustainable development, 34(1), 21-43.

Mu H., Zhou L., Dang X., Yuan B., 2019. Winter Wheat Yield Estimation from Multitemporal Remote Sensing Images based on Convolutional Neural Networks. 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), 5-7 Aug 2019.10.1109/Multi-Temp.2019.8866918

Narin, O. G., & Abdikan, S. (2020). Monitoring of phenological stage and yield estimation of sunflower plant using Sentinel-2 satellite images. Geocarto International, 1-15.

Nevavuori P., Narra N., Lipping T., Crop yield prediction with deep convolutional neural networks, Computers and Electronics in Agriculture, 163, 2019, 104859

Ouattara, B., Forkuor, G., Zoungrana, B. J., Dimobe, K., Danumah, J., Saley, B., & Tondoh, J. E. (2020). Crops monitoring and yield estimation using sentinel products in semi-arid smallholder irrigation schemes. International Journal of Remote Sensing, 41(17), 6527-6549.

Palatnik, R. R., & Roson, R. (2012). Climate change and agriculture in computable general equilibrium models: alternative modeling strategies and data needs. Climatic Change, 112(3), 1085-1100.

Putt, E.D. 1977. Early history of sunflower. p. 1–19. In A.A. Schneiter (ed.) Sunflower technology and production. ASA, CSSA, and SSSA. Madison, WI.

Rouse, J., Haas, R., Schell, J., Deering, D. (1974). Monitoring vegetation systems in the great plains with ERTS. In: Third ERTS Symposium, 10–14 December 1974, Washington, USA, Vol. 1; p. 48–62.

Sameen M.I., Pradhan B., Lee S., 2020. Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment, CATENA, 186, 104249

Seelan, S. K., Laguette, S., Casady, G. M., & Seielstad, G. A. (2003). Remote sensing applications for precision agriculture: A learning community approach. Remote sensing of environment, 88(1-2), 157-169.

Smith, J., 1987a. Close range photogrammetry for analyzing distressed trees. Photogrammetria, 42(1), 47-56.

Smith, J., 1987b. Economic printing of color orthophotos. Report KRL-01234, Kennedy Research Laboratories, Arlington, VA, USA.

Smith, J., 2000. Remote sensing to predict volcano outbursts. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XXVII-B1, 456-469.

Tilling, A.K., O'Leary, G.J., Ferwerda, J.G., Jones, S.D., Fitzgerald, G.J., Rodriguez, D., Belford, R., 2006. Remote sensing of nitrogen and water stress in wheat. Field Crops Res. 104, 77–85.

Yan X., Su X.G. 2009. Linear regression analysis: theory and computing. World Scientific, SingaporeYang Q., Shi L., Han J.,

Zha Y., Zhu P., 2019. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images, Field Crops Research, 235, 142-153.

Zhang, G., Wang, M. & Liu, K. 2019. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. Int J Disaster Risk Sci 10, 386–403. https://doi.org/10.1007/s13753-019-00233-1