Prediction of vegetation indices

1st Dias Bakhtiyarov
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
dbakhtiyarov@nu.edu.kz

2nd Selim Temizer
School of Engineering and Digital Sciences
Nazarbayev University
Nur-Sultan, Kazakhstan
selim.temizer@nu.edu.kz

***Abstract*—Prediction of vegetation indices on wheat fields in Kazakhstan.**

***Index Terms*—Vegetation index, NDVI, NDMI, GNDVI, ClGreen, LSTM, Sentinel-2, remote-sensing**

I. Introduction

On a global scale, Kazakhstan is considered as one of the main exporters of wheat in the world food market. This means that ability to predict the amount of yield produced at the end of the growing season can help government and financial bodies make important decisions earlier, in a more effective and efficient manner so as to reduce possible wheat price fluctuations and increase the income. Prediction of yield on a local scale is also important,since farmers will be able to better foresee the right amount of equipment and labor force required for successful harvesting of his/her fields. More importantly, information about the predicted yield can help farmers gain cause-and-effect relationship between the actions on the field and their impact on the final yield at the end of the growing season.

With the advent of remote sensing technologies, one of the promising approach towards quantification of the vegetation activity is considered to be visual images captured by satellites equipped with multi-spectral imaging (MSI) instrumentation. Sentinel-2 twin satellites managed by European Space Agency (ESA) are equipped with such MSI instruments and are capable of providing top-of-atmosphere imagery products with spatial resolution of up to 10m and temporal resolution of 5 days. These products are distributed open-source and are available for download through Copernicus Open Access Hub (COAH).

There are growing research interest on application of satellite-based remote sensing data for prediction of crop yields. However, direct prediction of crop yields is often not possible, since it requires presence of an abundant amount of field-collected (labeled) data. Another important fact to consider is that the yields for wheat crop are harvested only once per growing season (time period from April to September). Sentinel-2 satellites started its program in 2016, which implies that only 4 temporal samples (growing seasons in years 2016, 2017, 2018 and 2019) per field can be collected. Hence, reliance on field-collected crop yield data is not a feasible solution to date.

Instead of predicting the crop yield at the end of growing season, another idea is to predict vegetation activity of wheat crop throughout its growth period within the growing season between April and September. Such approach would mean much more amount of samples that can be collected for successful prediction of vegetation activity one or several time steps ahead.

As regards application of remote sensing technologies, there are multiple ways to quantify the conditions of vegetation on Earth surface. These include Normalized Difference Vegetation Index (NDVI), Normalized Difference Moisture Index (NDMI), Green Normalized Difference Vegetation Index (GNDVI) and Chlorophyll Index Green (ClGreen). These calculation of these indices are based on multi-spectral band information that can be obtained from Sentinel-2 satellite derived products.

The main objective of this study is, therefore, to predict the level of vegetation activity of wheat crop one month ahead based on historical spatio-temporal data of multi-spectral information obtained from Sentinel-2 satellite-based remote-sensing imagery data.

II. Dataset

The remote sensing dataset consists of 8000 imagery products corresponding to four years of time period between April and September within each year. Spatially, these images belong to three oblasts of northern region of Kazakhstan, which are Akmola, Kostanay and North Kazakhsan oblasts.

Most of the satellite images are available as top-of-atmosphere (TOA) satellite images. However, for more accurate prediction of the vegetation activity, it was necessary to convert TOA products to bottom-of-atmosphere (BOA) products by performing atmospheric correction using “sen2cor” software provided by COAH.

Each of obtained 8000 BOA products (single BOA takes up around 1GB of disk space) contains the multi-spectral bands of a 100x100km2 land region. The resolution of bands, as well as their spectral identification are presented on Fig. 1.

The ”sen2cor” software additionally performs image classification by providing scene classification (SCL) map, example of which is provided on Fig. 2. Using SCL map it was possible to remove bad pixels, regions of dense clouds and their shadows, as well as any other pixel regions containing non-vegetation objects, such as lakes, rivers, city zones, etc.

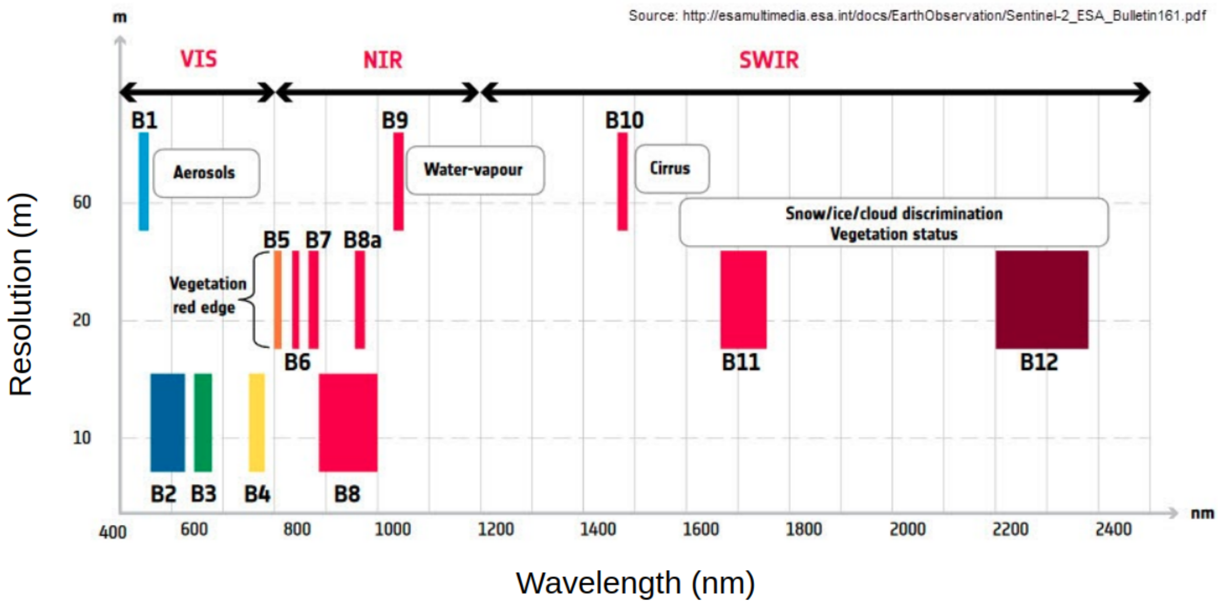

Fig. 1. Spectral composition of Sentinel-2 satellite based remote sensing product.

Each of the aforementioned BOA products are then cropped according to individual wheat field geometries. In three oblasts of northern Kazakhstan, namely Akmola, Kostanay and North Kazakhsan, 37298 fields were identified that correspond to one of the wheat crop species (either Triticum aestivum L or Triticum durum Desf). The cropping process consists of determining all the wheat fields that are located within a BOA product region (100x100km2) and then cropping each of the product bands (total 9 bands) along those field geometries. Fig. 3 provides a visual example of such cropping. All pixels corresponding to absence of useful data (blue regions), as well as those corresponding to non-vegetation pixel regions, are replaced with NaN values so as to leave only the meaningful data on each of the cropped bands.

The next step was to calculate vegetation indices, which are NDVI, GNDVI, NDM and ClGreen. The example of vegetation indices derived from a set of wheat fields are provided on Fig. 4. Due to limitation in time and space, it was decided to only record the mean values of the indices, thus focusing only on global temporal changes of the vegetation activity in fields.

The acquired vegetation indices are then scaled between 0 and 1 before feeding it into a model.

III. Model

Long Short Term Memory (LSTM) model is utilized for prediction of vegetation activity at wheat fields. The model architecture is provided on Fig. 5. The model is constructed in Python programming language using PyTorch library.

IV. Experiments and Results

Overall 89 experiments were performed with different hyperparameters for LSTM model, including size of hidden layers, number of hidden layers, learning rate and number of epochs. For each experiment, training loss and validation loss (val_loss) metrics were recorded. The full list of model traininig experiments and runs are provided in a separate file, wheareas Fig. 6 represents the top 10 runs that scored best on validation dataset. The training process is provided in Fig. 7.

V. Conclusion

Overall, using vegetaion activity as the attributes for the LSTM model, it was possible to obtain abundant amount of samples (about 210000 samples), which effectively helped to accurately train the model. The LSTM model proved to be effective at predicting vegetation activity of wheat fields. The project is available and can be accessed online at ”https://github.com/lgblkb/dm_final_project”.

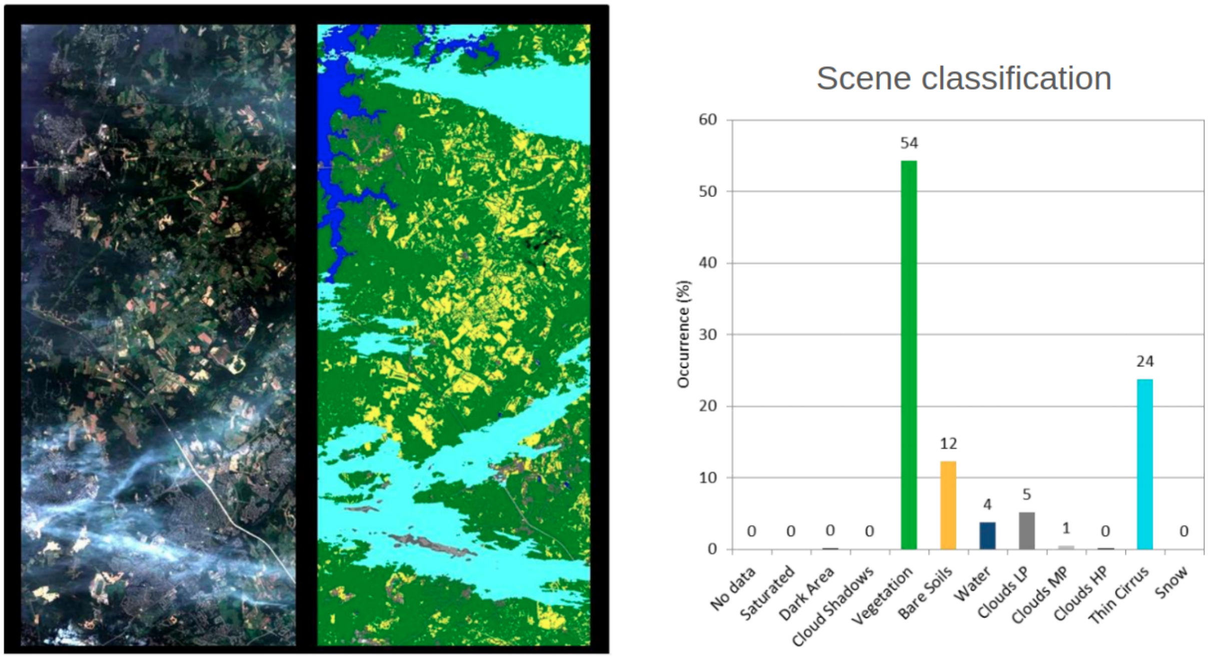

Fig. 2. Scene classification map obtained as a result of performing atmospheric correction using ”sen2cor” software.

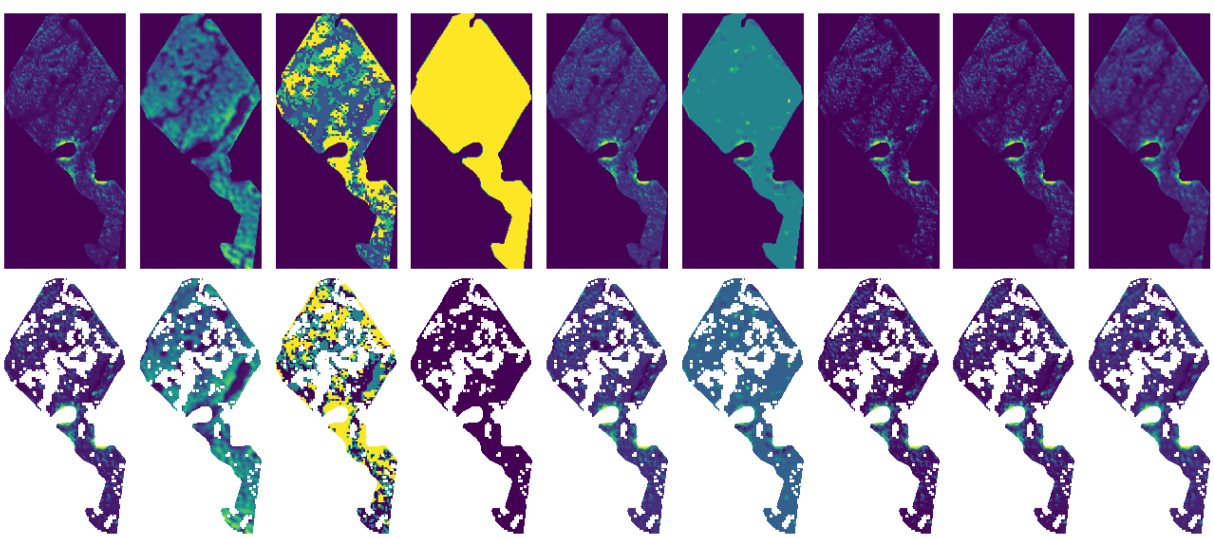

Fig. 3. The results of cropping each of the bands of the bottom-of-atmosphere imagery product. First row images: bands containing the field geometry. Second row images: bands cleared from non-vegetation indices.

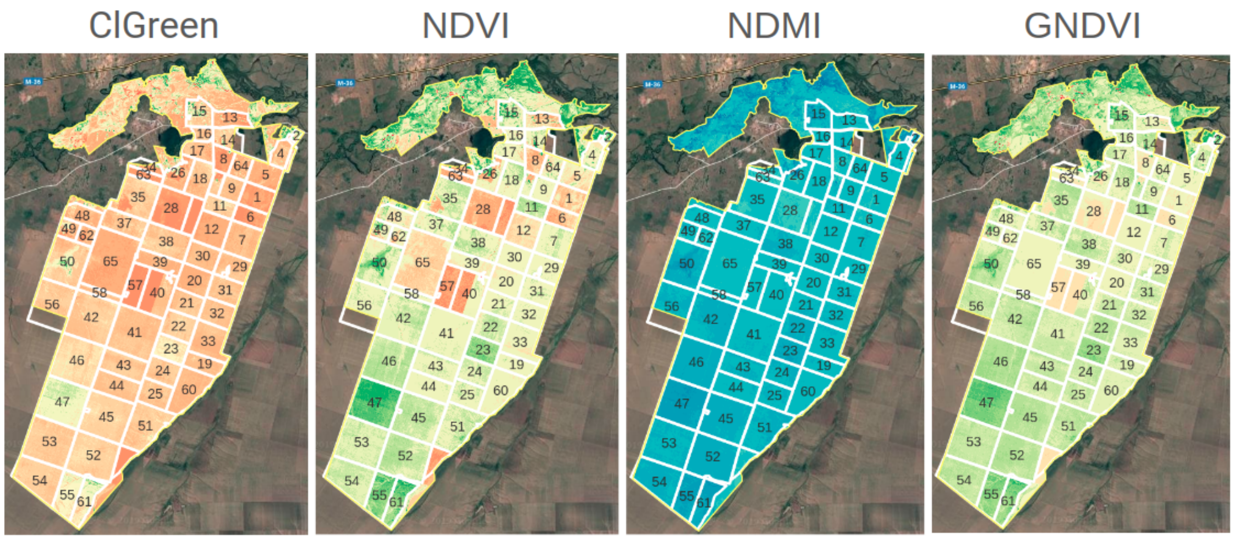

Fig. 4. Vegetation indices.

	Name	Type	# Parameters	Output Shape
●	lstm	LSTM(4, 128, num_layers=5, …	2048, 65536, 512, 512, 65536…	20000,4,128, 5,20000,128,5,…
●	fc	Linear(in_features=128, out…	512, 4	20000,4

Fig. 5. Vegetation indices.

Runtime	batch_size	hidden_size	learning_rate	num_epochs	num_layers	Training loss	val_loss ▲
2m 14s	20000	98	0.01179	43	3	0.002165	0.002107
2m 28s	20000	100	0.01157	44	4	0.00215	0.002121
2m 22s	20000	117	0.007789	47	5	0.00207	0.002134
1m 49s	20000	112	0.017	35	6	0.002151	0.002142
2m 5s	20000	124	0.009456	41	10	0.002181	0.002145
1m 54s	20000	151	0.01152	36	9	0.002182	0.002149
1m 56s	20000	170	0.01373	37	9	0.002213	0.002153
1m 58s	20000	145	0.009103	37	10	0.002179	0.00217
2m 1s	20000	108	0.01265	38	4	0.002209	0.00218
1m 39s	20000	115	0.01532	30	5	0.002214	0.002199

Fig. 6. Vegetation indices.

Fig. 7. Vegetation indices.