Data Article

Deep potato – The hyperspectral imagery of potato cultivation with reference agronomic measurements dataset: Towards potato physiological features modeling

Bogdan Ruszczak\textsuperscript{a,b,*}, Dominika Boguszewska-Mańkowska\textsuperscript{c}

\textsuperscript{a} QZ Solutions Sp. z o.o., Ozimska 72A St., Opole 45-310, Poland
\textsuperscript{b} Department of Electrical, Control and Computer Engineering, Opole University of Technology, Prószkowska 76 St., Opole 45-758, Poland
\textsuperscript{c} Plant Breeding and Acclimatization Institute - National Research Institute Jadwisin, Serock 05-140, Poland

\textbf{A R T I C L E   I N F O}

\textit{Article history:}
Received 3 December 2021
Revised 9 March 2022
Accepted 16 March 2022
Available online 21 March 2022

Dataset link: Deep Potato (Original data)

\textbf{Keywords:}
Hyperspectral imagery (HSI)
Potato agronomic measurements
Potato physiology
Image processing
Soil moisture
Deep learning

\textbf{A B S T R A C T}

The datasets provide hyperspectral imagery of potato fields with referencing agronomic measurements of several parameters. It contains also meteorological data collected on the place at the same time and some additional data on the variety of potatoes and the experiment. The experiment has been conducted in 2020 and two different potato varieties (Lady Claire and Markies) on the different soil profiles were planted and observed. During that time, on 4 different days, to provide a detailed picture of the experiment the hyperspectral imagery has been taken using a UAV and 150-band hyperspectral camera. The collected material has been later processed into 8 geo-referenced orthophotomaps. To provide precise reference information that could be later used for modeling purpose the measurements of plants from each field has been performed. The registered data contains data on plant height, number of stems, stem fresh and dry mass, leaf fresh and dry mass, leaf assimilation area and LAI.

\* Corresponding author at: Faculty of Electrical, Control and Computer Engineering, Opole University of Technology, Prószkowska 76 St., Opole 45–758, Poland.
\textit{E-mail address:} b.ruszczak@po.edu.pl (B. Ruszczak).

https://doi.org/10.1016/j.dib.2022.108087
2352-3409/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
number of tubers, tuber fresh and dry mass, starch content, RWC, chlorophyll a fluorescence index, the maximum quantum yield of PSII photochemistry, and the performance of electron flux.

© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

### Specifications Table

| Subject | Computer Science - Computer Vision and Pattern Recognition |
|---------|----------------------------------------------------------|
| Specific subject area | Hyperspectral imagery of potato fields with referencing laboratory measurements |
| Type of data | Hyperspectral imagery |
| How the data were acquired | The imagery was collected using a push-broom line imaging spectrometer with 150 spectral bands in the range of 460-902 nm, that was mounted on a UAV. The measurements using calibrated lab weight and scale, Mintola SPAD-502 chlorophyll meter, Pocket-PEA fluorimeter, LI-3100A area meter, laboratory dryer Sup-100, polarimeter Palamat S |
| Data format | Raw (measurements with locations - gejson, orthophotomaps - Jpeg2000) Enhanced (extracted individual binary images - NPY) |
| Description of data collection | The hyperspectral imagery has been processed into 8 georeferenced orthophotomaps that report measured reflectance registered at 80-meter flight level – 4 maps for each of 2 locations (each location had a different soil profile). Reference measurements, which were carried out by standard laboratory methods, provide the information on: |
| | - plant height [cm], |
| | - number of stems, stem fresh mass [g] and dry mass [g], |
| | - leaf fresh mass [g], and leaf dry mass [g], |
| | - leaf assimilation area, and LAI – measured using LI-3100A area meter (LI-COR, USA) instrument, |
| | - number of tubers, tuber fresh mass[g], and tuber dry mass [g], |
| | - starch content [%] using polarimeter Palamat S, |
| | - the degree of hydration of leaf tissue using RWC [%] (RWC column), |
| | - photosynthesis efficiency as chlorophyll a content using SPAD index using Mintola SPAD-502 device (SPAD column), |
| | - the maximum quantum yield of PSII photochemistry (Fv/Fm column) measured using Multifunctional Plant Efficiency Analyzer device (Handy-PEA fluorimeter, Hansatech Instruments Ltd., Pentney, King’s Lynn, Norfolk, England) and Pea Plus software, |
| | - Performance index: the performance of electron flux to the final PSI electron acceptors (Pi column) |
| | - potato variety;measurement location (precise location measured with RTK GPS device denoted as polygon pointing to the area where plant material was collected). |
| | - Lastly, the 84 individual hyperspectral images were extracted from the specific parts of orthophotomaps paired accordingly to measurements locations. Each extracted hyperspectral image depicts one field part where the plant’s material has been collected. |
| Data source location | • Plant Breeding and Acclimatization Institute - National Research Institute (IHAR-PIB) |
| | • Jadwisin, Legionowo County, Masovian Voivodeship, Poland |
| | • N 52° 29.00562 E 21° 2.67360 |
| Data accessibility | Repository name: Deep potato |
| Data identification number: DOI: | ⟨13:italic⟩10.17632/xn2wy75f8m.1⟨/13:italic⟩ |
| Direct link to the dataset: | https://data.mendeley.com/datasets/xn2wy75f8m |
Value of the Data

- This imagery dataset is aimed to aid the modeling of the measured plants’ parameters. The application of the hyperspectral camera results in more detailed information, separated into 150 spectral channels, which could provide information that is not visible while using regular cameras.
- Anyone interested in HSI imagery, for which the precise lab measurements are supplemented can benefit from these data.
- The set could also aid comparison of the different vegetation indices calculations, such as NDVI or LAI calculated using provided imagery to provide precise measurements (i.e. measured potato leaves assimilation surface or LAI).
- There are no other datasets publicly available for some registered plants parameters, especially sets of hyperspectral images with precise labels on RWC, chlorophyll a content (SPAD), the maximum quantum yield of PSI photochemistry, and the performance of electron flux.
- The plants' measurements provided in the set together with hyperspectral imagery could help to provide estimation machine learning or deep learning models that could estimate those parameters.

1. Data Description

- Folder “orthophoto_maps” contains 8 high resolution, hypersectral, 150 channel orthophotomaps. Files stored in ENVI format – two files for each map (HDR file with the necessary information and BSQ file with imagery data; additionally the BSQ files are compressed as ZIP): “2020-07-30_10_11_53_ihar_16_g1.hdr”, “2020-07-30_10_29_ihar_15_g2.hdr”, “2020-07-21_10_06_21_ihar_14_g1.hdr”, “2020-07-21_10_05_04_ihar_13_g2.hdr”, “2020-07-14_10_05_31_ihar_15_g2.hdr”, “2020-06-30_10_27_39_ihar_25_g2.hdr”, “2020-06-30_10_24_15_ihar_22_g1.hdr”, “2020-07-30_10_11_53_ihar_16_g1.zip”, “2020-07-30_10_29_ihar_15_g2.zip”, “2020-07-21_10_06_21_ihar_14_g1.zip”, “2020-07-21_10_05_04_ihar_13_g2.zip”, “2020-07-14_10_06_47_ihar_16_g1.hdr”, “2020-07-14_10_05_31_ihar_15_g2.hdr”, “2020-06-30_10_27_39_ihar_25_g2.hdr”, “2020-06-30_10_24_15_ihar_22_g1.hdr”, “2020-07-30_10_11_53_ihar_16_g1.zip”, “2020-07-30_10_29_ihar_15_g2.zip”, “2020-07-21_10_06_21_ihar_14_g1.zip”, “2020-07-21_10_05_04_ihar_13_g2.zip”, “2020-07-14_10_06_47_ihar_16_g1.zip”, “2020-07-14_10_05_31_ihar_15_g2.zip”, “2020-06-30_10_24_15_ihar_22_g1.zip”, “2020-06-30_10_24_15_ihar_22-g1.zip”
- “measurements/table1.geojson” – table with reference lab measurements, with precise geolocation,
- “measurements/table2.geojson” – table with reference physiological measurements, with precise geolocation,
- Folder: “images”, with files: “1.npz”, “2.npz”, “3.npz”, … “84.npz” - binary NPY files – individual hyperspectral image files extracted from orthophotomaps (name of each file is provided in the measurement Tables 1 and 2 accordingly to the column: "sample_index"); files could be opened using NUMPY package (https://numpy.org),
- dataset_presentation.html – a file with a brief presentation of the dataset

2. Experimental Design, Materials and Methods

The measurements of this dataset were made for potato cultivation in 2020 in IHAR-PIB. The experiment was carried out in natural conditions on 24 micro-plots. Each micro-plot is a separate object with an area of 5.6 m² with a soil profile fenced off with a layer of concrete to a depth of ten meters to avoid water leakage. The experiment included 2 soil profiles: heavy clay sand on medium clay – 12 plots and light loamy sand on light clay – 12 plots. Two potato varieties Lady Claire and Markies were planted (2020-04-28) on each soil profile (6 plots of each soil profile per 1 variety). 24 plants were planted on each micro-plot. The enumeration of the plots for the two locations (called: group 1 and group2) is presented in Fig. 1.

Plant protection treatments were applied in accordance with the principles of Good Agricultural Practice. One herbicide treatment, two treatments with insecticides (against the Colorado beetle), and four treatments with fungicides (against late blight and Alternaria) were performed.
Table 1
The average measurements for the second measurements geojson file concerning the planted potato variety and the soil profile (group).

| Parameter                  | Lady Claire Group 1 | Group 2 | Markies Group 1 | Group 2 |
|----------------------------|---------------------|---------|-----------------|---------|
| FvFm                       | 0.769               | 0.709   | 0.749           | 0.698   |
| PI                         | 1.897               | 0.896   | 1.316           | 0.658   |
| RWC                        | 85.250              | 85.311  | 86.528          | 87.028  |
| SPAD                       | 43.490              | 29.945  | 41.649          | 31.382  |
| Assimilation Area          | 9597.25             | 7967.59 | 8252.00         | 9155.00 |
| LAI                        | 3.840               | 3.188   | 3.301           | 3.662   |
| Leaf Dry Mass              | 9.273               | 9.363   | 10.523          | 10.698  |
| Leaf Fresh Mass            | 661.34              | 625.83  | 638.16          | 606.39  |
| Number Of Stems            | 5.333               | 6.917   | 4.333           | 4.222   |
| Number Of Tubers           | 12.667              | 15.333  | 4.056           | 6.500   |
| Plant Height               | 78.167              | 75.333  | 97.278          | 101.333 |
| Starch Content             | 12.078              | 13.198  | 10.311          | 9.729   |
| Stem Dry Mass              | 6.223               | 6.205   | 7.521           | 7.038   |
| Stem Mass                  | 460.37              | 372.52  | 648.57          | 655.94  |
| Tuber Dry Mass             | 18.578              | 19.796  | 17.166          | 16.766  |
| Tuber Fresh Mass           | 432.96              | 585.68  | 176.10          | 238.46  |

Table 2
Categorical parameters encoding, for both geojson measurements files.

| Parameter     | Values                                                                 |
|---------------|------------------------------------------------------------------------|
| Field         | 1, 2, 3, ..., 24;                                                      |
| Group         | 1: for twelve plots with light clay sand, 2: for twelve with heavy clay sand; |
| Variety       | 0: for Lady Claire, 1: for Markies                                     |
| Fold          | 0, 1, 2, 3, 4: for training subsets, -1: for validation subset        |
| Sample Index  | 1, 2, 3, 4, ..., 83, 84                                               |

The imagery has been taken using a UAV and hyperspectral camera on 2020-06-30, 2020-07-14, 2020-07-21, and 2020-07-30 when the plants were fully developed. The imagery was collected using a push-broom line imaging spectrometer with 150 spectral bands in the range 460–902 nm. The collected material has been later processed into 8 georeferenced orthophotomaps – 4 maps for each location. The final ground sampling distance (GSD) for the map is 2.2 cm and each map covers at least 400 m².

For someone interested more in hyperspectral data than orthophotomaps the dataset provides the extracted hyperspectral images. They are stored as 150 bands NumPy masked arrays (they could be opened using the open-source NumPy library - https://numpy.org). In Fig. 2 several 2-band color compositions of those images are shown. In Fig. 3 for the same exemplary images, the extracted spectral characteristics are depicted (curves are composed of averaged pixel values of the following spectral bands).

To provide precise reference information that could be later used for modeling purpose the measurements of plants from each field has been performed. The registered measurements contain data on RWC, SPAD index, the maximum quantum yield of PSII photochemistry (Fv/Fm) and the performance of electron flux (PI) (that are depicted in Fig. 4) plant height, number of stems, stem fresh mass and stem dry mass, leaf fresh and dry mass, leaf assimilation area and LAI, number of tubers, tuber fresh and dry mass, starch content (presented on Fig. 5). More on the agronomic standards and measurements methodology implemented in IHAR-PIB is described in the paper [1].

The experiment included 2 soil profiles: heavy clay sand on medium clay and light clay sand on light clay - 12 plots. Different soil profiles were selected to be able to influence irrigation and the response of plants grown on different soil types. Soil profiles differ in their sorption properties (water retention in the soil).
Environmental conditions, including weather conditions, have a huge impact on the development and plant yield. Extremely unfavorable weather conditions inhibit plant growth and harvest, therefore it is necessary to determine them. The agronomic parameters, listed in the manuscript, determine the tolerance of potato plants to soil drought, plant productivity and are necessary for the plants’ condition and yielding assessment in various cultivation systems [2,3].

For a better understanding of the agronomic process, the weather parameters present during the experiment were logged as well. The data on air humidity measured on 2 m above ground level couple of meters from depicted fields - “H200”, air temperature measured on 2 m - “T200”, ground temperature 5, 10, 20, and 50 cm below ground level (“Tg5”, “Tg10”, “Tg20”, “Tg50”), and soil moisture of each field measured using gravimetric method.
Fig. 2. Three exemplary extracted hyperspectral images (in rows) – for each image, four different compositions of selected bands are depicted: first column – composition of red, green, and blue bands, the second column – a composition with enhanced brightness, third column – image cropped using provided mask (each *.npz file in the dataset provides necessary mask data), and fourth column - CIR composition of near-infrared, red and band.

Fig. 3. Averaged pixel values of the following spectral bands of the example images. Each line corresponds to the images presented in Fig. 2.
Fig. 4. The distributions of the measured RWC, SPAD, FvFm and PI parameters (listed in: “Table 1.geojson”) with respect to the planted potato variety and the soil profile (Group).

Fig. 5. The distributions of the second set of measured parameters (listed in: “Table 2.geojson”) with respect to the planted potato variety and the soil profile (Group).
The measurements tables contain some additional useful information:
- "Flight_Date" (when the acquisition took place), "Map_Filename" – the name of the ortophotomap where the measurement is depicted, "geometry" – encoded location of the measurement,
- and some data describing the experiment: "Sample_Index" – number of measurement and number of extracted images, "Field" – number of observed micro-plot, "Group" – number of observed group of fields, "Variety" – potato variety, "Fold" – number of proposed subsets for modeling (where: -1 – test set, 0 to 4 – parts of the training set),
- the encoding of the additional categorical fields has been described in detail in the Table 2.

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Data Availability

Deep Potato (Original data) (Mendeley Data).

CRediT Author Statement

Bogdan Ruszczak: Conceptualization, Methodology, Software, Data curation, Visualization, Writing – original draft, Formal analysis; Dominika Boguszewska-Mańkowska: Methodology, Investigation, Validation, Formal analysis, Writing – review & editing.

Acknowledgments

Funding:
This work was partially supported by The National Centre for Research and Development of Poland under project POIR.04.01.04-00-0009/19.

References

[1] D. Boguszewska-Mańkowska, M. Pieczyński, A. Wyrzykowska, H.M. Klaji, L. Sieczko, Z. Szweykowska-Kulińska, B. Zagdańska, Divergent strategies displayed by potato (Solanum tuberosum L.) cultivars to cope with soil drought, J. Agro. Crop Sci. 204 (2018) 13–30, doi: 10.1111/jac.12245.
[2] D. Boguszewska-Mańkowska, B. Ruszczak, K. Zarzyńska, Classification of potato varieties drought stress tolerance using supervised learning, Appl. Sci. 12 (4) (2022) Issue, doi: 10.3390/app12041939.
[3] K. Zarzyńska, M. Pietraszko, Possibility to predict the yield of potatoes grown under two crop production systems on the basis of selected morphological and physiological plant development indicators, Plant Soil Environ. 63 (4) (2017) 165–170 issue, doi: 10.31721/161/2017-PSE.