Invariant learning as a pathway to robust potato yield prediction

Stelios P. Neophytides[a,c], Ilias Tsoumas[b,e], Andria Tsalakou[d], Michalakis Christoforou[d], Michalis Mavrovouniotis[a,c], Marinos Eliades[a,c], Christiana Papoutsa[a,c], Charalampos Kontoes[b], and Diofantos G. Hadjimitsis[a,c]

[a]ERATOSTHENES Centre of Excellence, 82 Franklinou Roosvelt, Limassol, Cyprus
[b]Beyond Centre of EO Research & Satellite Remote Sensing, National Observatory of Athens, 6 Karystou, Athens, Greece
[c]Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol, Cyprus
[d]Department of Agricultural Science, Biotechnology and Food Science, Cyprus University of Technology, Limassol, Cyprus
[e]Wagenigen University & Research, 4 Droevendaalsesteeg, Wagenigen, Netherlands

ABSTRACT

Yield prediction is an essential task to sustain the food market and to ensure the food for the world in the upcoming decades. Potatoes (Solanum tuberosum L.) are a vital staple food for many countries in the world and the advancement of accurate yield prediction will aid in promoting the agricultural industry. Potato is one of the most exportable agricultural products in Cyprus. Artificial Intelligence (AI) and Remote Sensing (RS) based agriculture monitoring has showed a massive impact in yield estimation in recent years. Monitoring vegetation indices like Normalized Difference Vegetation Index during the phenological stages of potatoes can provide identical insights into crop growth and yield. In this study, our focus lies on robust yield prediction across varied spatial and temporal dimensions. Specifically, we explore two distinct regions in Cyprus (i.e seaside and interior), each characterized by unique local agroclimatic conditions. The dataset encompasses potato yield data, in-situ meteorological data and vegetation indices derived by Sentinel-2 for a 7-years period (2017-2023). Thus, we test invariant learning against traditional ML methods in terms of spatial robustness and data drift issues.

Keywords: yield, food security, agriculture, remote sensing, causality, machine learning

1. INTRODUCTION

Potatoes are considered as the most widely known non-grain food crop globally. The high yield productivity, low water requirements required and the nutritional value makes potatoes a great contributor towards achieving world's food security.[1] The year 2008 was announced as the year of potato by the United Nations and leveraged potato's position as the most popular non-cereal food. Moreover potatoes are considered as the fourth largest food crop around the world after corn, rice and wheat. Nevertheless, there is a gap in the context of analyzing the environmental impact on potato's yield productivity.[2] The significance of potatoes also highlighted by Food and Agriculture Organization (FAO) as a valuable crop to vanish poverty and hunger.[3]

According to the official agricultural statistics report of the Republic of Cyprus, 550,096 tonnes of potatoes were harvested from 2014 to 2018. This is roughly a quarter of the total production of Cyprus' agricultural land. About 450.000 tonnes were exported to the rest of the world. The total amount of exported potatoes brought approximately 215 thousands euros to Cyprus' economy[*]. In contrast to other important crops for the agriculture

Further author information: (Send correspondence to Stelios P. Neophytides)
Stelios P. Neophytides: E-mail: stelios.neophytides@eratosthenes.org.cy
[*]https://data.gov.cy/index.php/el/resource/e-agriculture-1960-2018

industry of Cyprus and the economy of the island, potatoes are producing higher yield and brought a much more higher income from exports for Cyprus. Thus, it is quietly understandable that potatoes are considered as the most important economic and food crop of Cyprus.

Consequently, predicting potatoes' yield is significant for the future of food and agriculture sustainability worldwide. Remote sensing (RS) in the recent years plays a crucial role in agricultural monitoring.[4,5] One of the main aspects which concern RS as a source of information are yield prediction tasks for various crops.[6] Through the years, potato yield prediction evolved from ground-based RS techniques to aerial and satellite based RS with spectral indices (e.g., normalized difference vegetation index (NDVI), modified chlorophyll absorption ratio index (MCARI), leaf area index (LAI), etc.), structural, thermal and textural information derived from satellites playing a pivotal role in predictive tasks. Moreover, crop growth models and simulations, artificial intelligence methodologies and mechanistic roles have been utilized throughout the years for potato yield prediction.[7]

The most widely used vegetation indices in potato yield prediction tasks are NDVI, green NDVI (GNDI), normalized difference red-edge (NDRE), soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI), red-edge chlorophyll index($Cl_{red-edge}$) and ratio vegetation index (RVI).[8] More recent studies, showed the capabilities of RS in developing a new spectral indicator, the Potato Productivity Index using Sentinel-2 spectral information achieving a root mean square error of 15.42%.[9] In this study we are introducing two water (i.e., Normalized Difference Water Index and Normalized Difference Moisture Index) related spectral indices that as far as we know have not been considered as predictors for potato yield prediction in other studies.

A proposed approach couples WOrld FOod STudies (WOFOST) model and Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) model achieved an average difference of 1800 kilos per hectare using satellite-based LAI, total temperature statistics during different phenological stages of the plants and fraction of total dry matter among others.[10] Another study using various spectral indices from Sentinel-2 images and tested machine learning models showed that support vector machine radial achieved an accuracy of 11.16% root mean square error.[11] A research work conducted in Munshiganj district of Bangladesh tested NDVI from Landsat imagery and the regression analysis achieved a difference of 10.4% between actual and predicted values.[12] NDVI and other indices were used for regression analysis of potato's yield in Aroostook County, Maine, United States and achieved an R^2 of 0.12 to 0.41.[13] The fusion of data derived from unmanned aerial vehicle (UAV) and hyperspectral satellite imagery were fed in a Random Forest regression model combined with RReliefF feature selection algorithm and achievied a coefficient of determination higher than 0.9.[14] An approached based on Kriging (also known as Gaussian process regression) interpolation method achieved a R^2 of approximately 0.70 using NDVI acquired from Sentinel-2 in Tarakeswar, Hooghly, West Bengal, India.[15] Furthermore, a research work conducted with Belgium as the case study used Random Forest as the predictive algorithm and utilized NDVI, weather data and soil water depletion as the predictors and achieved a R^2 of 0.57 and 0.68 for late and early potato crops respectively.[16]

Most of the aforementioned studies are focused in wide areas and data are derived from different growing environments of potatoes were may differ in local agroclimatic conditions. Therefore, technologies like causal inference and causal machine learning[17] are useful to determine the causality of predictors and develop robust predictive models in different environments. This research work propose causal inference using invariant learning in Cyprus and discover underlying patterns among two different growing environments of potatoes. To the best of our knowledge, causal inference using invariant learning in agriculture and especially in yield prediction tasks has not yet widely investigated. In general, causal analysis and discovery through different techniques has been applied in different disciplines related to agriculture like pest management, effect estimations of farmers' practices and recommendations.[18–20]

Moreover, the aim of this work is to identify the causal predictors for potato yield prediction across two different growing environments in Cyprus towards developing more robust yield prediction models. The objectives of this research work are i) introduce new spectral indices in potato yield prediction tasks and ii) introduce causal inference using invariant learning in yield prediction tasks using earth observation and meteorological data. The proposed approach integrates in-situ meteorological data (i.e., air temperature, humidity, wind speed and rainfall) and spectral indices (i.e., NDVI, NDWI and NDMI) from Sentinel-2 data.

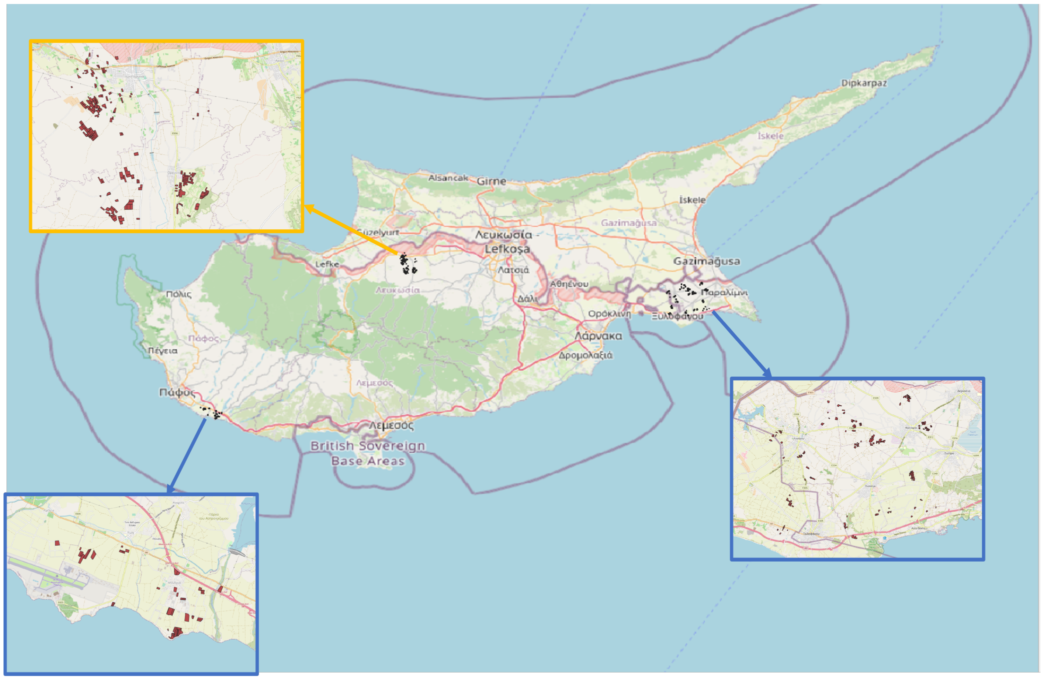

Figure 1. Map of potatoes' growing environments. Blue and yellow border indicates the seaside and interior environments respectively. Red polygons presents the agricultural parcels of potatoes.

The rest of the paper is structured as follows. Section 2 describes the materials used, the study area and the proposed methodology. Section 3 presents the experimental setup, the results of causal inference and discussions on the results. Major conclusions and future work is given in Section 4.

2. MATERIALS AND METHODS

2.1 Case study

Cyprus is located in the eastern Mediterranean Sea (35°N, 33°W) as shown in Figure 1, and is part of the spacious area of the Eastern Mediterranean, Middle East, and North Africa (EMMENA) region.[21] The island covers a total area of 9254 km^2 and the climate is identified as semi-arid with frequent drought events. In general, Cyprus experiences arid summers and humid winters which are typical characteristics of the Mediterranean region.[22,23] From 1991 to 2020, the average mean land surface temperature in Cyprus was 11.2 degrees $°C$ for the winter season, 16.82 $°C$ for the spring season, 26.69 $°C$ for the summer season, and 21.03 $°C$ for the autumn season, according to the reports of the Climate Change Knowledge Portal for 2021[†]. Winter season's precipitations amounts were 281.06mm, for spring were 90.9mm, for summer were 10.3mm, and for autumn were 87.81mm.

This study focuses on potatoes. Potatoes are cultivated in two different environments of the island. The first environment are the areas cultivated with potatoes near the coast (i.e., Xylofagou, Frenaros, Sotira, Liopetri, Kouklia, Timi and Mandria) and the other one is located in the interior area of Cyprus (i.e., Orounta and Peristerona). The two areas are characterized by different agroclimatic conditions, while also the cultivated potatoes near the sea are affected from the conditions (e.g.,moderate temperatures, sea spray, soil salinity, etc).

2.2 Yield data

The ground-truth data needed for this study were collected from farmers' associations in Cyprus. The data consists of parcel identification number based on the Department of Lands and Surveys of Cyprus, coverage area in decares, cultivar type, quantity of potato seed used during sowing, sowing date, harvesting date, paleo irrigation appliance and the harvested yield. The data are covering the period of 2019 to 2023. The collected

[†] https://climateknowledgeportal.worldbank.org/

Table 1. Descriptive statistics for the satellite and meteorological features used in this study. Temperature (T) is expressed in $^{\circ}C$, relative humidity (RH) is expressed as %, wind speed (WS) is expressed as $m \cdot s^{-1}$ and rainfall is expressed in mm. Yield is expressed in tonnes per decare. Min: minimum; Max: maximum; Mean: mean; stdev: standard deviation.

Feature	**Min**	**Max**	**Mean**	**stdev**
Yield	0.50	22.10	2.08	1.39
NDWI$_{median}$	-0.624	-0.103	-0.302	0.11
NDWI$_{min}$	-0.730	-0.161	-0.511	0.12
NDWI$_{max}$	-0.518	0.233	-0.106	0.10
NDWI$_{stdev}$	0.007	0.304	0.122	0.04
NDMI$_{median}$	-0.177	0.463	0.091	0.14
NDMI$_{min}$	-0.247	0.369	-0.092	0.07
NDMI$_{max}$	-0.098	0.669	0.357	0.13
NDMI$_{stdev}$	0.004	0.288	0.152	0.06
NDVI$_{median}$	0.038	0.740	0.280	0.17
NDVI$_{min}$	-0.124	0.622	0.068	0.08
NDVI$_{max}$	0.112	0.838	0.604	0.17
NDVI$_{stdev}$	0.000	0.352	0.184	0.07
Rainfall	0.00	482.60	209.40	111.75
T$_{max}$	19.30	44.50	33.11	5.40
T$_{min}$	-1.50	18.40	2.97	3.40
T$_{mean}$	12.00	29.15	16.98	3.58
RH$_{mean}$	59.00	80.00	70.59	4.07
RH$_{max}$	87.00	100.00	97.28	2.67
RH$_{min}$	7.00	40.00	17.01	5.04
WS$_{mean}$	1.60	4.10	2.78	0.34
WS$_{max}$	11.60	29.70	21.03	2.99

samples are geo-referenced according to the annual Land Parcel Identification System of Cyprus Agricultural Payments Organization. Geo-reference of samples, makes the acquisition of time-series data from satellite images, possible. After all the data cleaning required for this study, 809 agricultural parcels were available for the invariant learning.

2.3 Satellite-based vegetation indices

Satellite-based remote sensing collects data from various sensors in near-real time for a variety of environmental aspects, including vegetation, water, natural hazards, and topography. The uptake of such data sources let the agricultural industry meet significant improvements with large-scale and systematic monitoring of cultivated fields. In this work, data from the Sentinel-2 satellite of European Space Agency was used. Sentinel-2 constellations consists of two satellites and their joint revisit of an Earth's location happens every 5 days. The satellite is equipped with a multi-spectral instrument and captures information from the visible to the short-wave infrared spectrum. The spectral information is distinguished in 13 bands of different spatial resolutions (i.e., 10m, 20m and 60m).

In this study 1 vegetation and 2 water related indices are utilized as shown in Table 1. The Normalized Difference Vegetation Index (NDVI) is used as a proxy of the vegetation growth of potatoes cultivations. The index is calculated using the obtained surface reflectance of a target (i.e., a pixel) from a satellite at visible red and near-infrared wavelengths with the following expression:[?]

$$p_i = \frac{B8_i - B4_i}{B8_i + B4_i}, \tag{1}$$

where $B8_i$ and $B4_i$ represent the spectral reflectance at 842nm and at 665nm, respectively, for pixel i.

The Normalized Difference Moisture Index (NDMI) is used as a proxy of the vegetation water content of potatoes plants. The index is calculated using the obtained surface reflectance of a target (i.e., a pixel) from a satellite at near-infrared and short-wave infrared wavelengths with the following expression:[24]

$$p_i = \frac{B8_i - B11_i}{B8_i + B11_i}, \tag{2}$$

where $B8_i$ and $B11_i$ represent the spectral reflectance at 842nm and at 1610nm, respectively, for pixel i.

The Normalized Difference Water Index (NDWI) is used as a proxy of the wetted parcels either from precipitation events or irrigation appliance in a time near the observation. The index is calculated using the obtained surface reflectance of a target (i.e., a pixel) from a satellite at green and near-infrared wavelengths with the following expression:[25]

$$p_i = \frac{B3_i - B8_i}{B3_i + B8_i}, \tag{3}$$

where $B3_i$ and $B8_i$ represent the spectral reflectance at 560nm and at 842nm, respectively, for pixel i.

Potatoes in Cyprus are cultivated mainly in two different seasons of the year. Based on the sowing and harvesting dates, Sentinel-2 acquisitions were filtered. The data collection was done through Google Earth Engine. The mean value of each aforementioned index, on each agricultural parcel was calculated with the following abstract equation:

$$\overline{X^k} = \frac{1}{N^k} \sum_{i=0}^{N^k} p_i, \tag{4}$$

where $\overline{X^k}$ is the mean value of each index for each parcel k, N^k is the number of pixels within the boundaries of each agricultural parcel k and p_i are the index values at the ith pixel as defined in Equations (1), (2) and (3).

To better describe the growing season using the satellite variables acquired for each agricultural parcel, the minimum, maximum, median and standard deviation of each index were calculated throughout the growing season.

2.4 Meteorological variables

The meteorological variables , used in this study have been collected from the Department of Meteorology (DoM) of Cyprus. Despite that DoM has a dense meteorological network covering the whole island, a weather station is not available for each area cultivated with potatoes. Thus, for each area the meteorological variables from the nearest station are acquired. Each station is collecting every 10 minutes information about air temperature at 2m, wind speed, wind direction, relative humidity and accumulated rainfall as described in Table **??**. All of the data have passed from department's quality control and validation procedures. DoM is archiving the daily maximum, minimum and average air temperature at 2m, the daily maximum and average wind speed and the daily accumulated rainfall.

2.5 Causal inference using invariant predictions

The concept of causal inference using invariant predictions was first introduced by Jonas Peters on 2015 and later extended to non-linear models.[26,27] The concept investigates the invariance in causal relationships in different environments that took place in a problem to increase the robustness of predictive models. The environment can be defined as different areas, times, policies, or other contextual factors that affect the target variable. In this case, the two areas cultivated with potatoes, distinguished by their agroclimatic conditions, are considered as the two environments. The environment is considered as $e \in \epsilon$, with ϵ as the set of environments.

Invariant learning as a method assumes a linear model as follows:

$$Y^e = \mu + X^e \cdot \gamma^* + \epsilon^e\,, \tag{5}$$

where Y^e is the target variable in environment e, μ is the intercept term, X^e are the features affecting the target variable Y in environment e, γ^* are the coefficients associated with the features and ϵ^e is the error term.

As in all supervised problems we have a bundle of inputs which are considered as predictors, the same happens in invariant learning. The set of causal predictors which are described by the non-zero parameters of the γ^* is expressed as follows:

$$\hat{\mathbf{S}} := \{k \mid \gamma_k^* \neq 0\}, \tag{6}$$

where $\hat{\mathbf{S}}$ is the set of causal predictors, k is the set of inputs, γ_k^* are the coefficients of each input.

According to the above, below is the formalization of the concept.

For each subset S of $\hat{\mathbf{S}}$ and for each environment a linear regression is fitted from all the data across all of the environments as follows:

$$Y = X^S \cdot \hat{\beta}_{pred}(S) + \epsilon^e, \tag{7}$$

where Y is the target output, X^S is the subset of predictors in $\hat{\mathbf{S}}$, $\hat{\beta}_{pred}(S)$ represents the estimated coefficients and ϵ^e is the error term.

Then the residuals of the model are calculated with the expression below:

$$R = Y - X^S \cdot \hat{\beta}_{pred}(S), \tag{8}$$

where R are the residuals, Y is the observed value, X^S are the predictors used in the model and $\hat{\beta}_{pred}(S)$ are the estimated coefficients associated with the predictors.

Then two different null hypothesis tests are taking place. The first hypothesis is that the mean of the residuals R is the same across all environments. Here, I_e defines the residuals of a specific environment e. For this hypothesis a two-sample t-test is used. The t-test is performed between residuals I_e with the residuals of the remaining environments I. The performed tests on all environments are adjusted for multiple comparisons using Bonferroni corrections. The second hypothesis tests that the variance of residuals remains the same across all environments using F-test. As in the first hypothesis, the results are combined using Bonferroni corrections. Then the p-values from the two hypothesis tests are combined using twice the minimum value of the p-values obtained.

Subset S with combined p-values less than significance level α (outcome of Bonferroni corrections) are considered as a valid set of causal predictors in ϵ as follows:

$$\hat{S}(\epsilon) := \bigcap_{S:\mathrm{H}_0,S(\epsilon) \text{ not rejected}} S, \tag{9}$$

where $\hat{S}(\epsilon)$ is the set of invariant predictors across all the environments ϵ, S is the tested subset of predictors, H_0 is the null hypothesis and $S(\epsilon)$ is the not rejected set of causal predictors.

Otherwise, the subset S is rejected as follows:

$$\hat{\Gamma}_S(\epsilon) = \emptyset, \tag{10}$$

where $\hat{\Gamma}_S(\epsilon)$ is the conventional confidence interval for (1 - α) of the parameters $\hat{\beta}_{pred}(S)$.

Table 2. Results of causal inference using invariant predictions. R is the rainfall, T is the temperature, RH is the relative humidity and WS is the wind speed. The abbreviations min,max and stdev stands for minimum, maximum and standard deviation, respectively.

Satellite variables		Meteorological variables	
Feature	q-value	Feature	q-value
$NDWI_{min}$	0.00258961	R	0.00408014
$NDWI_{max}$	0.00278918	T_{max}	0.00439409
$NDWI_{stdev}$	0.00281443	T_{min}	0.00439409
$NDWI_{median}$	0.00277248	T_{mean}	0.00439409
$NDMI_{min}$	0.00281443	RH_{mean}	0.00439409
$NDMI_{max}$	0.00281443	RH_{max}	0.00373542
$NDMI_{stdev}$	0.00281443	RH_{min}	0.00439409
$NDMI_{median}$	0.00281443	WS_{mean}	0.00327645
$NDVI_{min}$	0.00281443	WS_{max}	0.00397956
$NDVI_{max}$	0.00258961		
$NDVI_{stdev}$	0.00281443		
$NDVI_{median}$	0.00281443		

3. EXPERIMENTAL RESULTS

3.1 Experimental setup

The dataset consists of 343 samples for the interior areas environment and 466 samples for the seaside areas environment. Data were cleaned by removing samples with extreme values in features after an extensive data discovery using data analytics techniques. Moreover, all the values were normalized with 0 as the minimum boundary and 1 as the maximum boundary. For the causal inference task, the dataset was separated in a dataset which consists of the meteorological variables and a dataset of the satellite variables. The two sets of variables are tested separately for two reasons: i) the technique used assumes a linear regression model and thus the separation reduces the complexity the linear model has to face and ii) the agroclimatic conditions (characterized by meteorological variables) differs in the two environments. The identified satellite and meteorological causal predictors combined can create a robust feature set.

3.2 Results and analysis

Table 2 presents the results after the causal analysis performed on the two sets of variables. The results are separated by a double line in the two sets. For each set the feature and the corresponding $q-$value are given. The $q-$value for each feature represents the false discovery rate in multiple hypothesis testing. The adjusted $p-$value for the set of satellite variables is 0.00281443 and for meteorological variables is 0.00439409. The $q-$value of $NDWI_{stdev}$, $NDMI_{min}$, $NDMI_{max}$, $NDMI_{stdev}$, $NDMI_{median}$, $NDVI_{min}$, $NDVI_{stdev}$ and $NDVI_{median}$ is equal to 0.00281443 and therefore equal to the p-value of the satellite variables set. The $q-$value for $NDWI_{min}$ is 0.00258961, for $NDWI_{max}$ is 0.00278918, for $NDWI_{median}$ is 0.00277248 and for $NDVI_{max}$ is 0.00258961 which are lower than $p-$value. T_{max}, T_{min}, T_{mean}, RH_{mean} and RH_{min} achieved a $q-$value equal to 0.00439409 which is identical to the $p-$value of the meteorological variables set.

The results suggests that all of the features in this set can be considered as causal predictors of potato yield in the two environments. In this case, features with lower q-values can be considered as stronger causal predictors, as in general, lower q-values indicates a higher confidence that a discovery is true. The features with low q-values from the two sets are fused to create a dataset with the causal predictors of higher confidence. The new datasets consists of $NDWI_{min}$, $NDWI_{max}$, $NDWI_{mean}$, $NDVI_{max}$, R_{total}, RH_{max}, WS_{avg} and WS_{max}.

The tuber formation of potatoes is in general prone to the volume of water content in soil, soil temperatures and other meteorological aspects of the environment during the different growth stages.[28,29] Invariant causal predictions showed that three $NDWI$ statistics (i.e., minimum, maximum and mean) which was used a proxy of the wetted area in the agricultural parcels. Moreover, the utilized approach showed that total rainfall amounts are also strong predictors in both of the environments. Both rainfall and $NDWI$ are related to water and this is total related to the vulnerability of tuber formation to highly moist soils.[30] Vegetation's greenness and density can differ when those are growing in different environments and thus the model shows a higher confidence of maximum $NDVI$ to be a causal predictor. Moreover, $NDVI$ is one of the most used index for potato yield prediction.[8,31] The levels of relative humidity must be between 50-80 % to enhance nutrient uptake, photosynthesis and support the physiological processes of the plants and thus the maximum RH levels of each environment can be valuable in a potato yield prediction model.[32] In addition, a selected feature is the WS. A previous study investigated the effect of wind speed in potatoes' cultivations as a part of evapotranspiration calculations.[33] Other studies showed the positive effect of windbreaks in increasing tuber yield by approximately 5-10%.[34]

4. CONCLUSIONS

In this work, causal invariant predictions were introduced towards developing more robust generalized predictive models for yield prediction tasks. Based on the experiment conducted in this study the following concluding remarks can be drawn. All the satellite and meteorological variables used in this study as predictors for potato yield can be considered as causal factors among the two growing environments (i.e., interior and seaside areas of Cyprus). Moreover, according to the calculated $q-$value for each of the predictors the following can be considered as stronger causal features: $NDWI_{min}$, $NDWI_{max}$, $NDWI_{median}$, $NDVI_{max}$, T_{max}, T_{min}, T_{mean}, RH_{mean} and RH_{min}.

Future directions of this work is the adaption of this methodology with non-linear models. Moreover, the evaluation of the resulted causal feature set will be implemented to understand their impact to the performance of predictive models. The growing period will also be considered as an "environment" in future studies. This methodology, can be widely used in different fields of Earth Observation in better adaptation for generalized models.

ACKNOWLEDGMENTS

The authors acknowledge the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (www.excelsior2020.eu). The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination and Development and the Cyprus University of Technology.

Sincere thanks to the Department of Meteorology of Cyprus and Cyprus Agricultural Payments Organization for supporting this research work with data provision.

REFERENCES

[1] Lutaladio, N. and Castaldi, L., "Potato: The hidden treasure," *Journal of Food Composition and Analysis* **22**(6), 491–493 (2009). International Year of the Potato.

[2] Economou, F., Papamichael, I., Voukkali, I., Loizia, P., Klontza, E., Lekkas, D. F., Vincenzo, N., Demetriou, G., Navarro-Pedreño, J., and Zorpas, A. A., "Life cycle assessment of potato production in insular communities under subtropical climatic conditions," *Case Studies in Chemical and Environmental Engineering* **8**, 100419 (2023).

[3] Wijesinha-Bettoni, R. and Mouillé, B., "The Contribution of Potatoes to Global Food Security, Nutrition and Healthy Diets," *American Journal of Potato Research* **96**, 139–149 (Apr. 2019).

[4] Karmakar, P., Teng, S. W., Murshed, M., Pang, S., Li, Y., and Lin, H., "Crop monitoring by multimodal remote sensing: A review," *Remote Sensing Applications: Society and Environment* **33**, 101093 (2024).

[5] Weiss, M., Jacob, F., and Duveiller, G., "Remote sensing for agricultural applications: A meta-review," *Remote Sensing of Environment* **236**, 111402 (2020).

[6] Ali, A. M., Abouelghar, M., Belal, A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E., Elwesemy, A., Kucher, D. E., Maginan, S., and Savin, I., "Crop yield prediction using multi sensors remote sensing (review article)," *The Egyptian Journal of Remote Sensing and Space Science* **25**(3), 711–716 (2022).

[7] Lin, Y., Li, S., Duan, S., Ye, Y., Li, B., Li, G., Lyv, D., Jin, L., Bian, C., and Liu, J., "Methodological evolution of potato yield prediction: a comprehensive review," *Frontiers in Plant Science* **14**, 1214006 (July 2023).

[8] Mukiibi, A., Machakaire, A. T. B., Franke, A. C., and Steyn, J. M., "A Systematic Review of Vegetation Indices for Potato Growth Monitoring and Tuber Yield Prediction from Remote Sensing," *Potato Research* (June 2024).

[9] Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L., "New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: a machine learning approach," *International Journal of Remote Sensing* **42**, 3426–3444 (May 2021).

[10] Ntakos, G., Prikaziuk, E., ten Den, T., Reidsma, P., Vilfan, N., van der Wal, T., and van der Tol, C., "Coupled wofost and scope model for remote sensing-based crop growth simulations," *Computers and Electronics in Agriculture* **225**, 109238 (2024).

[11] Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L., "Potato yield prediction using machine learning techniques and sentinel 2 data," *Remote Sensing* **11**(15) (2019).

[12] Newton, I. H., Tariqul Islam, A. F. M., Saiful Islam, A. K. M., Tarekul Islam, G. M., Tahsin, A., and Razzaque, S., "Yield Prediction Model for Potato Using Landsat Time Series Images Driven Vegetation Indices," *Remote Sensing in Earth Systems Sciences* **1**, 29–38 (June 2018).

[13] Zaeen, A. A., Sharma, L., Jasim, A., Bali, S., Buzza, A., and Alyokhin, A., "In-season potato yield prediction with active optical sensors," *Agrosystems, Geosciences & Environment* **3**(1), e20024 (2020).

[14] Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., Liu, J., and Jin, L., "Above-ground biomass estimation and yield prediction in potato by using uav-based rgb and hyperspectral imaging," *ISPRS Journal of Photogrammetry and Remote Sensing* **162**, 161–172 (2020).

[15] Singha, C. and Swain, K. C., "Evaluating the ndvi based rice and potato yield prediction map using gis geostatistical environment," in [*2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*], 1–7, IEEE (2022).

[16] Vannoppen, A. and Gobin, A., "Estimating yield from ndvi, weather data, and soil water depletion for sugar beet and potato in northern belgium," *Water* **14**(8), 1188 (2022).

[17] Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., and Silva, R., "Causal machine learning: A survey and open problems," (2022).

[18] Giannarakis, G., Tsoumas, I., Neophytides, S., Papoutsa, C., Kontoes, C., and Hadjimitsis, D., "Understanding the impacts of crop diversification in the context of climate change: a machine learning approach," (July 2023). arXiv:2307.08617 [cs, q-bio].

[19] Giannarakis, G., Sitokonstantinou, V., Lorilla, R. S., and Kontoes, C., "Personalizing sustainable agriculture with causal machine learning," (2022).

[20] Tsoumas, I., Sitokonstantinou, V., Giannarakis, G., Lampiri, E., Athanassiou, C., Camps-Valls, G., Kontoes, C., and Athanasiadis, I., "Causality and explainability for trustworthy integrated pest management," (2023).

[21] Eliades, M., Michaelides, S., Evagorou, E., Fotiou, K., Fragkos, K., Leventis, G., Theocharidis, C., Panagiotou, C. F., Mavrovouniotis, M., Neophytides, S., Papoutsa, C., Neocleous, K., Themistocleous, K., Anayiotos, A., Komodromos, G., Schreier, G., Kontoes, C., and Hadjimitsis, D., "Earth Observation in the EMMENA Region: Scoping Review of Current Applications and Knowledge Gaps," *Remote Sensing* **15**, 4202 (Aug. 2023).

[22] Giannakopoulos, C., Hadjinicolaou, P., Kostopoulou, E., Varotsos, K. V., and Zerefos, C., "Precipitation and temperature regime over Cyprus as a result of global climate change," *Advances in Geosciences* **23**, 17–24 (Feb. 2010).

[23] Hadjinicolaou, P., Giannakopoulos, C., Zerefos, C., Lange, M. A., Pashiardis, S., and Lelieveld, J., "Mid-21st century climate and weather extremes in Cyprus as projected by six regional climate models," *Regional Environmental Change* **11**, 441–457 (Sept. 2011).

[24] Rud, R., Cohen, Y., Alchanatis, V., Levi, A., Brikman, R., Shenderey, C., Heuer, B., Markovitch, T., Dar, Z., Rosen, C., et al., "Crop water stress index derived from multi-year ground and aerial thermal images as an indicator of potato water status," *Precision Agriculture* **15**, 273–289 (2014).

[25] cai Gao, B., "Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sensing of Environment* **58**(3), 257–266 (1996).

[26] Heinze-Deml, C., Peters, J., and Meinshausen, N., "Invariant Causal Prediction for Nonlinear Models," *Journal of Causal Inference* **6**(2), 20170016 (2018).

[27] Peters, J., Bühlmann, P., and Meinshausen, N., "Causal inference using invariant prediction: identification and confidence intervals," (2015).

[28] McGarry, A., Hole, C., Drew, R., and Parsons, N., "Internal damage in potato tubers: a critical review," *Postharvest Biology and Technology* **8**, 239–258 (Sept. 1996).

[29] Bethke, P. C., "Potato Tuber Lenticels: A Review of Their Development, Structure, Function, and Disease Susceptibility," *American Journal of Potato Research* **100**, 253–264 (Aug. 2023).

[30] Fiers, M., Edel-Hermann, V., Chatot, C., Le Hingrat, Y., Alabouvette, C., and Steinberg, C., "Potato soil-borne diseases. A review," *Agronomy for Sustainable Development* **32**, 93–132 (Jan. 2012).

[31] Piedallu, C., Chéret, V., Denux, J., Perez, V., Azcona, J., Seynave, I., and Gégout, J., "Soil and climate differently impact ndvi patterns according to the season and the stand type," *Science of The Total Environment* **651**, 2874–2885 (2019).

[32] Wheeler, R. M., Tibbitts, T. W., and Fitzpatrick, A. H., "Potato Growth in Response to Relative Humidity," *HortScience* **24**, 482–484 (June 1989).

[33] Tang, J., Wang, J., Wang, E., Yu, Q., Yin, H., He, D., and Pan, X., "Identifying key meteorological factors to yield variation of potato and the optimal planting date in the agro-pastoral ecotone in north china," *Agricultural and Forest Meteorology* **256-257**, 283–291 (2018).

[34] Wright, A. and Brooks, S. J., "Effect of windbreaks on potato production for the atherton tablelands of north queensland," *Australian journal of experimental agriculture* **42**(6), 797–807 (2002).