



Article Forecasting Rainfed Agricultural Production in Arid and Semi-Arid Lands Using Learning Machine Methods: A Case Study

Shahram Rezapour ¹^(b), Erfan Jooyandeh ²^(b), Mohsen Ramezanzade ³, Ali Mostafaeipour ^{2,4,5}^(b), Mehdi Jahangiri ⁶^(b), Alibek Issakhov ^{7,8}, Shahariar Chowdhury ^{4,5} and Kuaanan Techato ^{4,5,*}^(b)

- ¹ Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan; rezapourshahram@yahoo.ca
- ² Industrial Engineering Department, Yazd University, Yazd 89158-18411, Iran; erfan.joyandeh@gmail.com (E.J.); mostafaei@yazd.ac.ir (A.M.)
- ³ Department of Electrical and Computer Engineering, Noshirvani University of Technology, Babol 47148-71167, Iran; ramzanzadeh@nit.ac.ir
- ⁴ Faculty Environmental Management, Prince of Songkla University, Songkhla 90110, Thailand; mdshahariar.c@psu.ac.th
- ⁵ Environmental Assessment and Technology for Hazardous Waste Management Research Center, Faculty of Environmental Management, Prince of Songkla University, Songkhla 90110, Thailand
- ⁶ Mechanical Engineering Department, Shahrekord Branch, Islamic Azad University, Shahrekord 88156-73111, Iran; jahangiri.m@iaushk.ac.ir
- Department of Mathematical and Computer Modelling, Faculty of Mechanics and Mathematics, Al-Farabi Kazakh National University, Almaty 50040, Kazakhstan; Alibek.issakhov@kaznu.kz
- ⁸ Department of Mathematics and Cybernetics, Kazakh-British Technical University, Almaty 50000, Kazakhstan
- * Correspondence: kuaanan.t@psu.ac.th

Abstract: With the rising demand for food products and the direct impact of climate change on food production in many parts of the world, recent years have seen growing interest in the subject of food security and the role of rainfed farming in this area. Machine learning methods can be used to predict crop yield based on a combination of remote sensing data and data collected by ground weather stations. This paper argues that forecasting drylands farming yield can be reliable for management purpose under uncertain conditions using machine learning methods and remote sensing data and determines which indicators are most important in predicting the yield of chickpea. In this study, the yield of rainfed chickpea farms in 11 top chickpea producing counties in Kermanshah province, Iran, was predicted using three machine learning methods, namely support vector regression (SVR), random forest (RF), and K-nearest neighbors (KNN). To improve prediction accuracy, for each county, remote sensing data were overlaid by the satellite images of rainfed farms with a suitable slope and altitude for rainfed farming. An integrated database was created by combining weather data, remote sensing data, and chickpea yield statistics. The methods were evaluated using the leave-oneout cross-validation (LOOCV) technique and compared in terms of multiple measures. Given the sensitivity of rainfed chickpea yield to the time of data, the predictions were made in two scenarios: (1) using the averages of the data of all growing months, and (2) using the data of a combination of months. The results showed that RF provides more accurate yield predictions than other methods. The predictions of this method were 7-8% different from the statistics reported by the Statistical Center and the Ministry of Agriculture of Iran. It was found that for pre-harvest prediction of rainfed chickpea yield, using the data of the March-April period (the averages of two months) offers the best result in terms of the correlation coefficient for the relationship between the yield and the predictor indices.

Keywords: machine learning; rainfed agriculture; chickpea; random forest; support vector regression



Citation: Rezapour, S.; Jooyandeh, E.; Ramezanzade, M.; Mostafaeipour, A.; Jahangiri, M.; Issakhov, A.; Chowdhury, S.; Techato, K. Forecasting Rainfed Agricultural Production in Arid and Semi-Arid Lands Using Learning Machine Methods: A Case Study. *Sustainability* **2021**, *13*, 4607. https://doi.org/ 10.3390/su13094607

Academic Editor: Anastasios Michailidis

Received: 12 March 2021 Accepted: 16 April 2021 Published: 21 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

Agriculture is directly and indirectly the world's main source of food supply [1]. With the rising demand for food products due to global population growth, and also the direct impact of climate change on food production in many areas, food security is again becoming a crucial issue for many countries. The agricultural output of each region very much depends on its climate and the related parameters [2]. One of the most important determinants of agricultural production efficiency and output and consequently food security in a region is people's understanding of the region's climatic characteristics and their impact on the yield of different crops. This issue is especially important in rainfed farming. Rainfed farming is a type of agriculture in which crops are almost entirely irrigated by rainfall, which makes it extremely dependent on climatic conditions [3]. Legumes are a family of protein-rich agricultural products that are not only essential for a healthy diet, but also serve as an important source of income for millions of families who grow them alongside other crops. Farming legumes is also known to improve soil nitrogen levels, which plays an important role in sustainable production [4]. Chickpea (Cicer arietinum L.) is a drought-resistant plant that grows well in arid and semi-arid areas. This plant can tolerate not only high temperatures but also low temperatures to some extent [5], which makes it a very good option for rainfed farms. With a total output of 271,487 tons (2017), Iran is the world's ninth-largest producer of chickpea after the United States [6]. Almost 80% of Iranian chickpeas are produced in four provinces: Kermanshah, Lorestan, Kurdistan, and West Azerbaijan. Accounting for nearly 25.2% of the country's total chickpea production, Kermanshah province is the largest producer of chickpea in Iran. More than 80% of Kermanshah's chickpeas are produced in rainfed farms, which have an average yield of 456kg/hectare [7]. Kermanshah has a semi-arid mountainous climate [8], which is perfect for rainfed chickpea farming. However, in recent years, the province has struggled with water scarcity issues due to the effects of climate change, especially prolonged dry spells and droughts, and the consequent overexploitation of groundwater resources [9]. Allowing agriculture to expand beyond the region's ecological and environmental capacities has adverse consequences—such as soil erosion, desertification, land pollution, and environmental degradation—which will ultimately result in the destruction of the region's natural resources and the decline of its agricultural output; trends that will actually move the region away from sustainable development [10]. Any information about the vegetation cover of a region and how it is impacted by climatic variables can substantially benefit the management of agricultural yield in that region [11]. However, the traditional methods of gathering information about climate and vegetation cover could be not only time-consuming but also inaccurate. Today, such information can be collected more easily by using remote sensing technologies to monitor vegetation and climate variables and track their temporal and spatial changes on regional and global scales. Indeed, these technologies can offer an expansive and integrated view of any area with extremely high accuracy. By combining the remote sensing data with climatic measurements, one can create a real-time map of crop characteristics, which can be used in the quantification of annual net outputs on different scales and also the differentiation of vegetation covers on continental and regional scales, which could be highly valuable for effective land management [12,13]. Given the massive volume of these data (spatial and temporal) which grow with every second, many machine learning methods have been developed to take advantage of emerging computer hardware and their computing power in processing these data. Using the aforementioned methods and data, it is possible to predict the yield of different crops in different areas in order to guide government policies that prioritize food security, especially the decisions on the export and import of food products, and also boost the agricultural economy by helping farmers and agriculture-related businesses make more informed decisions about crops.

2. Review of Literature

Numerous studies have shown that massive spatio-temporal data obtained from the maps and satellite images produced by satellite spectroscopy can be combined with regional meteorological data to accurately predict crop yield on different scales using statistical models [11,14–18]. The global problems during the recent decades have forced many researchers to pay more attention to clean energy sources for economic growth and environmental issues [19–21], therefore using rainfed agriculture would help to reduce pollutions. To predict crop yield, climate data such as temperature, precipitation, sunshine hours, humidity, etc. must be given to these models as input data. While climate data very well describe the environmental conditions that affect crop growth, they are not the only parameters that affect crop yield and cannot fully represent growth conditions. Climate data are more useful in crop yield estimations when combined with plant growth data [17]. Satellite images taken with remote sensing tools, such as spectral images in different bands, can offer direct information about the growth condition of plants. These spectral bands include optical spectra (visible and infrared) and microwave and thermal bands. Among these, light spectra such as infrared and red are widely used in the monitoring of plant growth parameters. Given the nonlinear nature of the empirical equations for the relationship between crop yield, crop growth indices, and climatic variables and the massive volume of spatio-temporal data that must be processed in statistical models to predict crop yield, these predictions require extensive computation. However, such computations can now be done easily thanks to advanced computer technology. In recent years, researchers have shown growing interest in the use of machine learning methods to predict crop yield [22]. Support vector machine (SVM), random forest (RF), and K-nearest neighbors (KNN) are three of the machine learning methods that have been successfully used to predict crop yield through classification and regression analysis of remote sensing data and climatic measurements.

For example, Gandhi et al. [10] used an SVM-based method to predict rice crop yield in order to facilitate crop selection in 27 districts of Maharashtra state in India.

In a study by Medar et al. [14], support vector regression (SVR) was used to predict the sugarcane yield in Karnataka, India. These researchers used a combination of climate and soil data with the satellite maps pertaining to the 2008–2018 period as the remote sensing data.

Chen et al. [15] used SVM to predict the yield of paddy rice in Chongqing, China, based on the weather data of 34 stations in this region for the 1985–2012 period. After compared the results of SVM with those of back-propagation neural network (BPNN) and multiple linear regression (MLR), this study reported that SVM is a better method for determining the effects of climate variables on crop yield.

In a study by Ejaz et al. [13], three methods including artificial neural network (ANN), linear regression (LR), and SVR were used to predict wheat yield in Punjab, Pakistan. These researchers used fertilizer and pesticide data as well as weather and soil data. All of these data, which were related to a 21-year period, were collected from the office of statistics and agriculture in Sargodha.

Kouadio et al. [16] used an ANN to predict the yield of robusta coffee in Vietnam based on SOM data and then compared the results with the results of RF and MLR. The data used in this study included soil fertility and soil yield data randomly collected from 44 farms in five districts of Lam Dong Province and the monthly averages of weather data for a 30-year period.

In a study by Kim & Lee [17], corn yield in Iowa, USA was estimated by the use of SVM, RF, ERT, and DL. The data used in this study were extracted from MODIS images of the Terra satellite and SM data obtained from ESA, CCI, which offer comprehensive global information with a resolution of 0.25° from active and passive microwave sensors. For modeling, these researchers defined 13 time periods consisting of 5 months in which the crop is grown and harvested.

Narasimhamurthy and Kumar [18] used RF to predict the rice yield in Andhra Pradesh, India. They used three types of data including temperature, precipitation, and perception data to estimate the rice yield in the studied area. The accuracy of their predictions with RF was 85.89%.

Parviz [23] used two methods, SVR and MLR, to predict barley yield in three Iranian provinces with arid, semi-arid, and humid climates. The results of this study showed that SVR is a better method for making a prediction for the considered 36-year period.

In a study by Kuwata and Shibasaki [24], SVR and DL were used to predict corn yield in Illinois, USA. In this study, DL was said to be preferred because of its advantage in the automatic selection of important features and also its greater flexibility in dealing with predefined features.

Tiwari and Shukla [25] conducted a similar study with BPNN and compared their training times and errors. These researchers used precipitation data and growth indices for yield prediction.

Han et al. [22] used 8 machine learning algorithms to predict the yield of winter wheat in 629 districts of a province of China. For this purpose, they defined multiple time windows and introduced an integrated framework of weather, remote sensing, and soil data for yield prediction based on the Google Earth Engine. These researchers used three groups of weather data and remote sensing data and soil data for their predictions.

In a study by Sharifi [26], which aimed to predict barley yield in Boshruyeh, Iran, this objective was pursued in three stages: (1) building a framework for predicting barley yield based on five-year weather data such as temperature and precipitation and remote sensing data such as NDVI And EVI for 24 districts; (2) identifying the best machine learning method for predicting barley yield by comparing four methods, namely GPR, KNN, DT, BPNN, which GPR to be the best method; and (3) identifying the best time for predicting crop yield before harvest. Tables 1 and 2 compares the past studies on the prediction of crop yield with machine learning methods.

Sharifi [23] investigated prediction of barley yield in Iran, which is one of the rainfed crops that can be grown in the region, NDVI, EVI, temperature and evaporation indicators were considered. Although, other factors such as soil moisture and rainfall can be considered in the growth of rain-fed barley. The irrigation method based on ground water, which means that farmers use water pumps to pump water was also investigated. Studies such as [13,17,20] have considered crops that are grown by irrigation. There have been other studies related to crops that are mainly located in areas where there is a lot of rain, such [10,14–16,18,21]. Hen et al. [22] studies winter wheat, criteria such as air pressure that have slight changes in the growing season and can be considered almost constant. In irrigation agriculture, due to the fact that water resources are available and there is no vital need for rainfall, indicators such as rainfall can be ignored, such as [10]. In another study conducted in Iran for barley, rainfall index, temperature, and speed wind were considered as indicators affecting plant growth. It was found that wind speed index except in stormy areas, has no effect on forecasting and can be ignored.

SVR has been widely used in related studies such as [10,13–15,17,19,20,22]. The SVR that is an optimized model of the SVM method, can be used to predict problems with continuous data like what we use in this study. Other researchers showed that, according to the intrinsic characteristics of crops and plant growth conditions, this method is suitable for forecasting [13,14,20].

Studies that used RF method, showed that this method has an acceptable performance for predicting the yield of agricultural products in different conditions with different indicator [10,16,17,22]. Kim and Lee [17] specifically demonstrated the capability of this method in using remote sensing indicators. Due to the unstable nature of climate phenomena and factors affecting plant growth, using the KNN method along with other methods gives more reliable performance to the results [22].

n	A d	Date of	Location of Study	Cron	T 1						М	ethods						
KOW	Autnors	Research	(Country)	Стор	10015	RF	MLR	E-BPNN	DL	SVM	ERT	DT	NN	KNN	GPR	BST	BGT	LR
1	Gandhi et al. [10]	2016	India	Rice	WEKA-EXCEl					*								
2	Ejaz and Abbasi [13]	2018	Pakistan	Wheat	Phyton-Weka					*			*					*
3	Medar et al. [14]	2019	India	Sugarcane	SciKit-Python					*								
4	Chen et al. [15]	2015	China	Paddy rice	MATLAB		*	*		*								
5	Kouadio et al. [16]	2018	Vietnam	Robusta coffee	MATLAB	*	*						*					
6	Kim and Lee [17]	2016	United States	Corn	R software	*			*	*	*							
7	Narasimhamurthy and Kumar [18]	2017	India	Rice	R software	*												
8	Parviz [19]	2020	Iran	Barley			*			*								
9	Kuwata and Shibasaki [20]	2015	United States	Corn	SciKit-Python & CAFFE				*	*								
10	Tiwari and Shukla [21]	2020	India		MATLAB			*										
11	Han et al. [22]	2020	China	Winter wheat	MATLAB-WEKA- GEE	*				*		*	*	*	*	*	*	
12	Sharifi [26]	2020	Iran	Barley	GEE			*				*		*	*			
13	Present research	2020	Iran	Chickpea	SciKit-Python	*				*				*				

Table 1. Comparison of the past studies on the prediction of crop yield with machine learning methods.

The * symbol in each row indicates which methods (columns) in each research has been used and names of method ab-breviations explained as follow: BPNN: Back-Propagation Neural Networks, RF: Random Forest, MLR: Multiple Linear Regression, DL: Deep Learning, SVM: Support Vector Machine, ERT: Extremely Randomized Trees, DT: Decision tree, NN: neural network, KNN: K-nearest neighbor, GPR: Gaussian Process Regression, BST: Boost Trees, BGT: Bagging Tree, LR: Logistic Regression.

				Soil F	eatures							Clima	te Data							Satelli	te Data			Sam	pling Features
Row	Authors	Fertilizer	Organic Carbon	Soil Temperature	Hd	Soil moisture	Soil Mixtures	Rainy Days	Precipitation	Evaporation	Humidity	Sunshine Hours	Solar Radiation	Temperature	Mean Wind Speed	Air or Vapor Pressure	Crop Evapotranspiration	IVUN	EVI	LAI	FPAR	GPP	ET	Number of Samples	Sampling Interval
1	Gandhi et al. [10]								*					*			*							27	1998*-2002
2	Ejaz and Abbasi [13]				*				*		*			*											21 years
3	Medar et al. [14]			*		*			*		*	*		*			*							2	2008-2018
4	Chen et al. [15]							*	*		*	*		*										34	1985-2012
5	Kouadio et al. [16]	*			*				*				*	*										44	2013-2014
6	Kim and Lee [17]					*			*					*		*		*	*	*	*	*	*	94	2004-2014
7	Narasimhamurthy and Kumar [18]								*					*										13	2005-2015
8	Parviz [19]								*					*	*	*								3	1982-2017
9	Kuwata and Shibasaki [20]								*	-				~		-		*	-					90	2001-2010
10	Han of al [22]		*			*	*		*					*				*	*					629	2001_2014
12	Sharifi [26]									*				*				*	*					24	2015-2014
13	Present research					*			*		*			*				*	*	*	*	*	*	11	2010-2017

The is companyed of the past stated of the adda of the same sing for the stated of	Table 2. Comparison of the	past studies based on	the data types and	sampling features.
---	----------------------------	-----------------------	--------------------	--------------------

The * symbol in each row indicates which data sources (columns) in each research has been used. Data source abbrevia-tions explained as follow: NDVI: Normalized difference vegetation index, EVI: Enhanced Vegetation Index, LAI: Leaf Area Index, FPAR: Fraction of Photosynthetically Active Radiation, GPP: Gross Primary Production, ET: Evapotranspi-ration.

In this study, for the first time the prediction of rainfed chickpea farming for 11 counties with the highest chickpea production in Kermanshah Province of Iran was performed by the use of SVM, RF, and KNN machine learning methods, and a combination of remote sensing and local synoptic station data.

The above comparison shows that most of the past studies have been focused on crops that are either irrigated or cultivated in rainy areas. In Iran, however, rainfed agriculture in areas with limited access to water accounts for a fairly large portion of the country's total agricultural output and therefore plays a critical role in the food security of this country [3,6]. Considering the climate of Kermanshah and its evolving water scarcity issues, there seems to be some room for the expansion of rainfed farming in this province. Also, since this province has an international border (with Iraq) and good access to the region's major transit routes, it may be able to play a bolder role in the food security of the region, which may also generate an economic boost for its own farmers and agriculture-related businesses. In this study, we tried to predict the yield of rainfed chickpea farming in 11 counties with the highest chickpea production in Kermanshah by the use of SVM, RF, and KNN machine learning methods and a combination of remote sensing, weather, and soil data.

3. Geographical Characteristics of Kermanshah Province

Kermanshah is one of the western provinces of Iran and has a border with Iraq. This province consists of 14 counties and its capital is the city of Kermanshah. Kermanshah province has an area of 24,640 km² (17th largest in Iran), a semi-arid mountainous climate [8], an average altitude of 1400 m (above sea level), an average temperature of 14 °C, and average annual precipitation of 450 mm, which is of Mediterranean type [27]. The geographical location of Kermanshah province is shown in Figure 1. The province has a very diverse climate (for example, while Qasr-e Shirin in the west of the province has a hot climate, Paveh in the north of the province has a cold climate, and Kermanshah in the center of the province has a temperate climate). Since time immemorial, rainfed farming has accounted for a major portion of agricultural output in this area. Kermanshah has long been one of Iran's top five provinces in terms of chickpea production and currently accounts for 20% of total chickpea production in this country. Over the last 12 years, the average chickpea yield in this province has been 441 kg/hectare. In terms of geographical coordinates, Kermanshah province is extended between $45^{\circ}20'39''$ and $48^{\circ}01'58'''$ eastern longitudes and $33^{\circ}37'08''$ and $35^{\circ}17'08'''$ northern latitudes.



Figure 1. Study area.

4. Methodology

In this section, we first describe the factors affecting the growth of rainfed plants. The methods used to predict is then explained. First, 9 remote sensing indicators and its sources and features will be explained. Then, using three filters of slope, elevation, and dryland farm density, we consider only the data related to lands that have appropriate slope, elevation and in which dryland cultivation takes place. After that the remote sensing data are combined with seven meteorological data that obtained from Iran Meteorological Organization. Finally, by adding the actual amount of dryland chickpea yields, our main data set is created for each year. In the following, the machine learning methods that used this data set for forecasting are explained. After that validation technique that used to estimate the accuracy of these methods are explained.

4.1. Crop Yield Parameters

In this study, crop yield predictions were made based on two groups of data: remote sensing data collected by satellite and weather data collected by regional synoptic stations. The remote sensing data included the data from MODIS, which provide information about plant growth during planting, germination, growth, and harvest, and also the soil moisture and CHIRPS data, which offer information about the water used for irrigation in rainfed agriculture. The meteorological data, which were intended to give information about the environmental conditions of the plants, were obtained from ground stations. All collected data are presented in Appendices A and B.

4.1.1. Indicators of Plant Growth in the Region

MODIS indices that could be related to plant growth and crop yield—including gross primary production (GPP), enhanced vegetation index (EVI), leaf area index (LAI), fraction of photosynthetically active radiation (FPAR), normalized difference vegetation index (NDVI), and evapotranspiration (ET)—were obtained from the Terra satellite. Figure 2a-f show the averages of these induces for 11 counties in the study area for the period 2010– 2017. It should be explained that the counties in the middle part of the province have a temperate climate, those in the northern part have a cold climate, and those in the western part have a hot and dry climate. NDVI, the map of which is illustrated in Figure 2a, is a simple graphical index developed for remote sensing analysis to assess the presence or absence of vegetation in an area. This index is computed based on the difference between near-infrared and red-light spectra. NDVI values close to zero (between -0.1 and +0.1) usually indicate bare rock, sand, or snow surfaces. Higher positive NDVI values (about +0.2 to +0.4) indicate shrub and grassland cover. EVI has been developed to cover some of the limitations of NDVI and particularly to minimize atmospheric effects and differences in blue and red reflectance [28]. The EVI map of the area is shown in Figure 2b. This index varies in the range of -1 to +1. LAI, which is plotted in Figure 2c, is the fraction of the surface area of the vegetation canopy to the surface area of the ground it covers. To obtain LAI for farmlands, the leaf area of the plants in one square meter of land should be measured. Another key parameter for the analysis of plant structures in ecosystems and also in crop yield models is FPAR, which provides a lot of information about photosynthesis, energy exchange, and carbon exchange of vegetation covers [29]. The FPAR map of the study area is shown in Figure 2d. FPAR is the fraction of photosynthetically active radiation (light available for photosynthesis) that is absorbed by the plant canopy. ET, the map of which is shown in Figure 2e, is the sum of evaporation and transpiration. This index has been defined because the evaporation from the moist soil surface cannot be easily separated from the transpiration from the plant surface. It has been shown ET has a relationship with crop yield and could be effective in its estimation [30]. The GPP map of the area is shown in Figure 2f. GPP is the ratio of solar energy captured in the plant's sugar molecules during the photosynthesis process, which then plants use for their cellular metabolisms, such as respiration and growth. Soil moisture content is also a key parameter for plant growth. Figure 2g-h shows the surface soil moisture and subsurface soil moisture, which

both have an impact on planting, germination, harvesting of rainfed plants. In general, soil moisture is the water content in the soil, which can be expressed in terms of mass and volume percentages. In rainfed farming, where crops are not irrigated by any water source other than rain, soil moisture is an important determinant of crop yield, although it is also affected by a wide range of variables. In this study, the soil moisture map of the study area was created using NASA-USDA SMAP Global soil moisture data.





Figure 2. Satellite images of the indices obtained from the remote sensing data.

The sources of the data used in the study and the temporal and spatial resolutions of the collected remote sensing data are given in Table 3. The indices related to plant growth were all obtained from MODIS data of the Terra satellite. The specifications of MODIS data products used in the study are also listed in the table. Given the differences in temporal and spatial resolutions, the satellite images were constructed using the monthly averages of the data.

Table 3. Specifications of remote sensing data used in the study.

Data	Spatial Resolution	Temporal Resolution	Source	Product
Normalized difference vegetation index	250 m	16-Day	NASA	MOD13Q1.006
Enhanced vegetation index	250 m	16-Day	NASA	MOD13Q1.006
Leaf area index	500 m	4-Day	NASA	MCD15A3H.006
Fraction of photosynthetically Active radiation	500 m	4-Day	NASA	MCD15A3H.006
Gross primary production	500 m	8-Day	NASA	MOD17A2H.006
Evapotranspiration	500 m	8-Day	NASA	MOD16A2.006
Rainfed croplands	1000 m	2010	NASA	USGS/GFSAD1000_V1
Surface soil moisture	0.25°	31-days	NASA	NASA_USDA/HSL/SMAP
Subsurface soil moisture	0.25°	31-days	NASA	NASA_USDA/HSL/SMAP
Precipitation	0.05°	Daily	CHIRPS	chirps-v2.0.1981-2019.39yrs.tif.gz

4.1.2. Indicators of Environmental Conditions of Plant Growth

To obtain the factors of environmental conditions of plant growth, the daily weather data collected by 11 synoptic stations of the Iran Meteorological Organization in the study area over an 8-year period were gathered. This included eight types of data, namely minimum temperature, maximum temperature, average temperature, sunny hours, maximum humidity, minimum humidity, average humidity, and dry temperature in growing, germination, and flowering seasons. Environmental factors also included slope and altitude, which were determined using the DEM of Kermanshah province. After giving a score to each 100 m² parcel of land, this map was transformed into the raster format. Since multiple layers of data had to be combined, the data of the layers were standardized before overlay. As the slope map illustrated in Figure 3a shows, a large portion of the province has a slope of 0–10 degrees. The northwestern and northeastern parts of the province generally have slopes of more than 10 degrees. In slopes of over 10 degrees, the shallowness of the soil and poor plowing conditions make the land more susceptible to environmental degradation. Between the mountains of the province, there are several plains with very gentle slopes from northwest to southeast. Figure 3b shows the altitude map of the study area. As can be seen, in Kermanshah province, altitude decreases from east to west. The conditions are more suitable in the central plains of the province, where the average altitude ranges from 800 to 1500 m. Some east and northeastern parts of the province, such as Songor County, have an average altitude of over 1500 m. In contrast, the western parts of the



province—such as Qasr-e Shirin, Sarpol-e Zahab, and Gilan-e Gharb counties have an average altitude of less than 800 m.

(a) Slope map of the study area

(**b**) Altitude map of the study area

Figure 3. Altitude and slope Satellite images of the study area.

4.1.3. CHIRPS Data

The rainfall data of the study area were collected from the monthly rainfall data contained in the CHIRPS dataset. The monthly cumulative rainfall data of CHIRPS have been obtained by the integration of global monthly precipitation climatology (CHPclim) data, precipitation estimates based on the TIR band of satellites, and data collected from ground stations. This dataset covers the geographical coordinates ranging from -50 to 50 degrees and time frames ranging from 1981 to the present. Several previous studies have shown the good consistency of this dataset with the data of ground stations [31]. According to [32], the average correlation of this dataset with Iran's rainfall data for spring, summer, autumn, and winter months is 0.61, 0.56, 0.59, and 0.45, respectively. CHIRPS dataset was downloaded from the website of the University of California, Santa Barbara (https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_annual/tifs (accessed on 24 October 2020)). To create the rainfall map shown in Figure 4, we used the 8-year average of daily rainfall data of Kermanshah. This map shows that the annual rainfall is highest in the northern counties such as Paveh and lowest in the western counties such as Qasr-e Shirin. The ArcPy library was used to perform mathematical operations on daily rainfall raster files in millimeters.



Figure 4. Raster map of annual rainfall in the study area based on CHIRPS dataset.

4.1.4. Rainfed Chickpea Farming and Yield Statistics of the Area

The rainfed chickpea production statistics of the study area were obtained from the database of the Statistical Center of Iran and the annual reports of the Ministry of Agriculture of Iran at the county level. Since the location information of rainfed farms by crop type was not available, we used GFSAD products to create raster images of rainfed farms in the study area. Combining the data of multiple satellite sensors, these products allow users to determine the type of land in terms of irrigation. Figure 5 shows the map of rainfed farms in Kermanshah according to GFSAD products (2010) with a resolution of 1 km.



Figure 5. Raster map of rainfed farms in Kermanshah province.

4.2. Machine Learning (ML) Methods

Machine learning (ML) is a method for data analysis that automates construction of model. It is a kind of AI (artificial intelligence) based on the theory that computers which can learn from data, identify patterns and make decisions without being programmed to perform specific tasks and automatically improve by experience and by implementing of data. ML learn from past calculations for producing repeatable decisions, reliable, and results. The repetitive aspect of machine learning is essential because the models are exposed to new data, also are able to independently modify patterns according to the new data. An essence aim of a ML learner is to generalize from its experience. Generalization is the ability of a learning machine to perform accurately on unseen samples after taking a learning data set. The training samples come from some generally unknown probability distribution and the learner has to construct a general model about this space that empowers ML learner to produce appropriately accurate predictions in new cases. Depending on the nature of the feedback which the ML learner receives from its environment, ML methods are usually divided into three categories; supervised, unsupervised, and reinforcement learning. In this study, we use supervised learning methods. Supervised learning methods build a mathematical model of a data set that includes both the inputs and its related real outputs—also known as a supervisory feedback—and learn a function that can be used to predict the output associated with new inputs. The data set is known as training data, and consists of a set of training samples. Here, we mean data set of remote sensing and meteorological data which are integrated and described in Section 4.1. An optimal function will allow the ML method to accurately predict the output for inputs that were not a part of the training data set. When the algorithm can do the prediction well and improves the accuracy of its predictions is said learned to perform that task. Supervised methods consist of two classification and regression algorithms, classification algorithms are used when the outputs are restricted to a limited set of classes, for example, sorting emails to spam and unspam emails. Regression algorithms are employed when the outputs may have any numerical value within a range, for example, what we do in this study prediction of crop yield according to growing conditions data set. The machine learning methods used in this study for yield chickpea prediction were random forest (RF), support vector machine (SVM), and K-nearest neighbors (KNN). First, these methods were trained with a set of training data, and then their ability to predict new data was evaluated by evaluation methods which are explained in Section 4.3.

4.2.1. Support Vector Machine

As in many other machine learning methods, SVM's modeling process involves two stages of training and testing. At the end of the training phase, testing data must be used to evaluate the generalizability of the trained model [33,34]. The solution of the SVR problem, which is the regression model of SVM, on a dataset consisting of L samples of form $\{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L), x \in \mathbb{R}^m, y \in \mathbb{R}\}$ is a linear function in the form of Equation (1), which can estimate output values based on inputs [33]

$$\hat{y} = \langle w.x \rangle + b \tag{1}$$

In this equation, \hat{y} is the estimated value, *x* is the input vector, w is the weight vector, and b is the bias.

As shown in Equation (2), the cost function introduced by Vapnik ignores the errors within a certain distance of the true value (epsilon). In other words, it considers some deviations to be acceptable [33]. The control parameters of the response function, i.e., weight and bias vectors, are obtained by solving the optimization problem of Equation (3) [35]

$$L(y,\hat{y}) = \begin{cases} 0 & if \quad |y - \hat{y}| < \varepsilon \\ |y - \hat{y}| - \varepsilon & otherwise \end{cases}$$
(2)

In this equation, ε is the acceptable error in the cost function, \hat{y} is the estimated value, and y is the true value.

$$Minimize: L(w,\xi) = \frac{1}{2} ||w||^2 + C \sum_{i=1}^{N} (\xi_i^* + \xi_i)$$

subject to:
$$\begin{cases} wx_i + b - y_i \le \varepsilon + \xi_i^* & i = 1, 2, ..., N \\ y - (wx_i + b) \le \varepsilon - \xi_i & i = 1, 2, ..., N \\ \xi_i^+ \ge 0, \xi_i \ge 0 & i = 1, 2, ..., N \end{cases}$$
(3)

In this equation, $||w||^2$ is the weight vector, ξ_i^* and ξ_i are the slack variables, *C* is the penalty or tuning parameter.

Using Lagrange coefficients, the optimization problem of Equation (3) can be converted to the Lagrangian function of Equation (4). The Lagrangian coefficients α and α^* are obtained by maximizing this function under the defined constraints.

$$L(\alpha^*, \alpha) = -\varepsilon \sum_{i=1}^{l} (\alpha_i^*, \alpha_i) + \sum_{i=1}^{l} y_i(\alpha_i^*, \alpha_i) - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\alpha_i^*, \alpha_i) \left(\alpha_j^*, \alpha_j\right) (x_i \cdot x_j)$$

$$subject \ to: \qquad \begin{cases} \sum \alpha_i^* = \sum \alpha_i \\ 0 \le \alpha_i^* \le C \ i = 1, 2, \dots, l \\ 0 \le \alpha_i \le C \ i = 1, 2, \dots, l \end{cases}$$

$$(4)$$

The above optimization problem can be solved with the help of quadratic programming methods, in which case the global extremum will be certainly reached and there will be no risk of being trapped in local extremums [35]. Thus, after computing α and α^* in Equation (4), the optimal weight vector w_0 , the optimal bias b_0 , and the prediction function of Equation (1) can be written as Equations (5)–(7) [36]

$$w_0 = \sum_{support \ vectors} (\alpha_i^*, \alpha_i) x_i \tag{5}$$

$$b_0 = -\frac{1}{2}w_0.[x_r + x_s] \tag{6}$$

$$f(x) = \sum_{support \ vectors} (\alpha_i^*, \alpha_i)(x_i.x) + b_0$$
(7)

In these equations x_i is the input vector with which the model is trained, χ is the new input vector, xr and xs are support vectors, w_0 is the optimal weight vector, and b_0 is the optimal bias value.

Data points with non-zero Lagrangian coefficients are called support vectors. Geometrically, these data points have a larger prediction error than epsilon. The parameters C and ε should be specified by the user. C is a tuning parameter and can be set to values ranging from zero to infinity. If C is set to large values, SVR will not allow errors in the training data, which results in lower generalizability. If C is set to values closer to zero, SVR will allow larger errors, because then the value of slack variables in Equation (3) becomes less important [36,37] and the model becomes less sensitive to the occurrence of errors in the training dataset. The parameter ε can also take values from zero to infinity. The value of this parameter greatly affects the condition of support vectors and consequently the performance of the model. Although setting ε to very large values will decrease the number of support vectors, which in itself is desirable, it is wrong to pursue this goal by widening the ε -band. On the other hand, setting ε to very small values will result in having a large number of support vectors, which increases the risk of over-training [37].

4.2.2. Random Forest

This method operates based on a structure consisting of numerous regression trees (CARTs) and its output is computed by integrating (or averaging) the output of these CARTs. The principles of ensemble training techniques are based on the assumption that some training methods perform better in certain situations and therefore using a group of these methods can result in good performance in a wide range of situations. In other words, a combination of several models (in our case, prediction models) tends to be more accurate than a single model. A decision tree is a non-parametric statistical model (in the sense that no value is pre-set for its variables) first introduced by Breiman et al. [38,39], which can be described as a structure consisting of many nodes and leaves that grow in the training process. To train a tree, one needs a set of training data like $L_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ where *n* is the number of observations, *x* is the input vector with m features $X = \{x_1, x_2, \dots, x_m\}$, and y is the output scalar. During training, the input data are split at each node, beginning with the root node. Each node uses its own splitting operator for this purpose. This process continues until the largest possible number of leaves are obtained. At the end of this training process, the prediction function $t = (X, L_n)$ is built on the training data. In the random forest (RF) method, each tree creates a random sample of data (hence the term random in the name). All of the generated decision trees are then combined by a bootstrap-aggregation (bagging) algorithm [40]. A bootstrap dataset is a set of instances like $(L_n^{\theta_1}, L_n^{\theta_2}, \dots, L_n^{\theta_q})$ that are obtained by random sampling with replacement from the training data. In RÉ, each decision tree is trained with a set of these values. Using the q trees created for the new input, the method obtains q predictions in the form of $\hat{y}_1 = t(x, L_n^{\theta_1}), \hat{y}_2 = t(x, L_n^{\theta_2}), \dots, \hat{y}_q = t(x, L_n^{\theta_q})$, and then integrates the predictions of individually trained trees by Equation (8) for regression problems [41,42].

$$\hat{y} = \frac{1}{q} \sum_{i=1}^{q} \hat{y}_i = \frac{1}{q} \sum_{i=1}^{q} t\left(x, L_n^{\theta_i}\right)$$
(8)

In this equation, \hat{y}_i is the prediction of the i-th tree.

The main advantage of using a bagging algorithm is that it helps avoid correlation between different trees, which results in the higher diversity of trees created from different samples. Another advantage of this algorithm is its noise resistance, because it produces independent trees based on different training samples, and while a decision tree alone could be noise-sensitive, the combination of several independent decision trees tends to be less noise-sensitive [43].

4.2.3. K-Nearest Neighbors

First introduced by Cover & Hart [44], KNN is a machine learning method used for regression. In this application, KNN considers each data record as a vector in an m-dimensional space (where m is the number of features), and predicts the value of each new sample based on the values of K records that are closest to that point in that space [45]. In this method, the closeness of the new point *x* and the training point x_i is measured by a Euclidean distance function in the form of Equation (9) as following [45]

$$d(x, x_i) = \sqrt{\sum_{j=1}^{m} \left(x_i^j - x^j\right)^2}, \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m$$
(9)

In this equation, *n* is the number of training samples and m is the number of input samples.

The steps of KNN can be summarized as follows.

Step 1: Specifying the parameter K (the number of nearest neighbors).

Step 2: Computing the distance of the new input from all training samples.

Step 3: Sorting the computed distances in ascending order and selecting the K samples with the shortest distance.

Step 4: Forming the set of values of the selected K samples.

Step 5: Computing the predicted value of the new sample by weighted averaging of the values in the set of Step 4.

The first step in this method is to choose the value of K (the number of neighbors). After selecting K, the Euclidean distance between the new sample and the existing samples, i.e., $d(x, x_i)$ for i = 1, 2, ..., n must be computed. The values of the K samples with the lowest d (shortest distance from the new sample) must then be put in Equation (10) to compute the predicted value of the new input [46]

$$\hat{y} = \sum_{i=1}^{k} w_i y_i \tag{10}$$

Here, w_i is the weight of each training sample, which depends on its distance from the new sample and is given by Equation (11)

$$w_{i} = \frac{\frac{1}{d(x,x_{i})}}{\sum_{i=1}^{k} \frac{1}{d(x,x_{i})}}$$
(11)

As this equation shows, samples that are closer to the new sample will have a greater impact on the prediction. In KNN, the new values are produced based on the previous data regardless of how much noise they have, which makes KNN sensitive to noise and irrelevant features [46].

4.3. Cross Validation

Overfitting is a modeling error that happens when a function is too closely fit to a limited data set. In such a case, the model tries to cover the data from the sample and even the noise values with a lot of changes. While such a model should reflect the behavior of society. In such cases, if obtained regression model is used to predict another sample, the predicted values will not seem appropriate at all. We used cross validation to overcome the problem of overfitting of ML methods. Cross-validation is model validation technique for assessing how the results of a mathematical analysis will generalize to unseen data set. Cross-validation randomly split the training dataset into *k* equal subsets. Then train MI method on k-1 of them and measure the prediction error on the *k*th subset, and do this in such a way which all *k* subsets be used for training and for calculating the prediction error. Then get average of the prediction errors from all *k* subset. Leave-one-out cross-validation

(LOOCV) is a kind of cross-validation which number of folds will be equal to number of instances in data set. Thus, the learning algorithm is applied once for each instance, using all other instances as a training set and using the selected instance as a single-item test set. Good machine learning models such as SVR, RF, and KNN are the ones which have good prediction accuracy, in other words, small prediction error shows that the method is capable of accurately predicting when faced with unfamiliar data. Validation measures quantify the deviation between the predicted data and the true data [17]. In this study, the accuracy of the machine learning methods for predicting the yield of products was evaluated using the LOOCV technique with MAE, MBE, RMSE, MAPE, and CC as measures. The equations of these measures are given below [15,17,26].

Mean absolute error (MAE) is a measure of errors between paired observations, which refers to the real yield of chickpeas obtained from the Statistical Center of Iran and the value predicted by forecasting methods. MAE is calculated as Equation (12) [15,17,26]

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$
(12)

It is thus an arithmetic average of the absolute errors. In these equations, y_i denotes the true values, \hat{y}_i denotes the predicted values. The mean absolute error uses the same scale as the data being measured. Mean bias error (MBE) is mainly used to estimate the average bias in the model and to determine if any steps need to be taken to correct the model bias. MBE captures the average bias in the forecasting. MBE is calculated as Equation (13) [15,17,26]

$$MBE = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$
(13)

In these equations, y_i denotes the true values, \hat{y}_i denotes the predicted values and n is the number of samples. RMSE is a measure of the differences between predicted values and real values or the quadratic mean of these differences. The RMSE serves to aggregate the errors in predictions for various data points into a single measure. In other word, is a measure of accuracy, to compare forecasting errors of different models for a particular data set. RMSE is calculated as Equation (14) [15,17,26]

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}}$$
(14)

where, y_i denotes the true values, \hat{y}_i denotes the predicted values and n is the number of samples. Mean absolute percentage error (MAPE) is also a measure of prediction accuracy of a forecasting method and is commonly used as a loss function for regression problems and in model evaluation, because of its very intuitive interpretation. This method expresses the prediction error as a percentage and is easier to express. MAPE is calculated as Equation (15) [15,17,26]

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{(\hat{y}_i - y_i)}{y_i} \right|}{n} \times 100$$
(15)

Here, y_i denotes the true values, \hat{y}_i denotes the predicted values, and n is the number of samples. Correlation coefficient (CC) is a statistical relationship between two variables including real and predicted values of chickpea yield. CC assumes values in the range from -1 to +1, where 0 the strongest possible disagreement and ± 1 indicates the strongest possible agreement. A number closer to 1 indicates a direct relationship between the two input arguments. Conversely, a number closer to -1 indicates an inverse relationship between them. Correlation coefficient is calculated as Equation (16) [15,17,26]

$$CC = \frac{\sum_{i=1}^{n} (\hat{y}_{i} - \overline{\hat{y}}_{i})(y_{i} - \overline{y}_{i})}{\sqrt{\sum_{i=1}^{n} (\hat{y}_{i} - \overline{\hat{y}}_{i})^{2} \sum_{i=1}^{n} (y_{i} - \overline{y}_{i})^{2}}}$$
(16)

In these equations, y_i denotes the true values, \hat{y}_i denotes the predicted values, and $\overline{\hat{y}}_i$ and \overline{y}_i are their respective averages and n is the number of samples.

Clearly, there are similar and other methods of prediction which could be implemented for any research works [47–50].

5. Results and Discussion

Since goal of this research was to estimate the yield of rainfed chickpea farms in Kermanshah, information about other types of land and other areas were irrelevant. Therefore, we combined the raster map of rainfed farms of the province with the raster maps of slope and altitude (Figure 3a,b) in order to filter the obtained satellite data and reduce irrelevant and outlier data. Since the crop yield data provided in the reports of the Statistical Center of Iran and the Ministry of Agriculture of Iran are at the county level, other data were also transformed into county level by the use of zonal operators. For this purpose, all pixels except those related to rainfed farms, slopes of less than 10 degrees, and non-mountainous heights were filtered out. In this way, we created an integrated database of remote sensing data, weather data, and rainfed chickpea farms in Kermanshah. Rainfed spring chickpea, which accounts for the largest portion of chickpea production in Kermanshah, is typically planted in late winter and harvested in late spring. Since the selected factors had different degrees of sensitivity to the time of data, we defined two scenarios for using the data of each month and also the combination of consecutive months (e.g., MA for March and April): (1) scenario (A), in which we used the average values obtained for the entire cultivation period from planting to harvest; (2) scenario (OC), in which we used an optimal combination of months depending on to the correlation coefficient for the relationship between the crop yield and each predictor variable. Table 4 shows the correlation coefficient for the relationship between the average value of each factor (predictor) in each month between March and June (planting, germination, growth, and harvest months for rainfed chickpea farming in the area) and the crop yield in 2010–2017. The highlighted cells in this table show the periods selected for the OC scenario on account of having the highest correlation coefficient. Table 5 illustrate the average of real and predicted values by the methods for the whole province in different years from 2010 to 2017 according to these two scenarios.

Table 4. Correlation coefficient of the relationship between the yield of spring chickpea and the values of its predictors during the growing months (from planting to harvest).

Criteria	March	April	May	June	MA	AM	MJ
Temperature min	0.059	0.009	0.028	0.081	0.033	0.019	0.056
Temperature max	0.079	0.093	0.094	0.087	0.086	0.094	0.091
Average Temperature	0.009	-0.018	-0.019	-0.018	-0.005	-0.018	-0.018
Precipitation	0.072	0.01	-0.366	-0.459	0.05	-0.124	-0.377
Humidity max	0.162	0.176	0.139	0.039	0.175	0.155	0.091
Humidity min	-0.317	-0.152	-0.013	0.044	-0.227	-0.080	0.008
Humidity	0.207	0.205	0.163	0.087	0.213	0.183	0.132
Dry temperature	0.103	0.211	0.294	0.238	0.151	0.261	0.269
Normalized difference vegetation index	0.458	0.503	0.142	0.068	0.616	0.291	0.113
Enhanced vegetation index	0.361	0.549	0.141	0.082	0.624	0.312	0.120
Fraction of photosynthetically active radiation	0.415	0.457	0.132	0.044	0.547	0.281	0.104
Leaf area index	0.462	0.456	0.148	0.044	0.559	0.291	0.118
Gross primary production	0.373	0.407	0.132	-0.005	0.566	0.260	0.084
Evapotranspiration	-0.021	0.235	0.134	0.041	0.161	0.180	0.110
Surface soil moisture	0.105	0.090	0.097	0.075	0.099	0.093	0.095
Subsurface soil moisture	0.203	0.202	0.203	0.221	0.204	0.204	0.210

Highlighted cells are the maximum in each row.

Year	Real Value of Production Kg/ha	Prediction by RF(OC) Kg/ha	Prediction by SVR(OC) Kg/ha	Prediction by KNN(OC) Kg/ha	Prediction by RF(A) Kg/ha	Prediction by SVR(A) Kg/ha	Prediction by KNN(A) Kg/ha
2017	542.76	535.27	509.31	524.56	514.13	495.45	484.85
2016	539.82	534.15	519.13	521.92	528.545	516.89	522.44
2015	281.85	320.31	368.36	395.32	341.79	392.87	413.74
2014	568.83	525.92	486.75	462.44	524.11	453.12	447.45
2013	475.52	454.20	446.63	424.78	457.32	428.88	446.95
2012	470.49	442.52	449.05	432.69	449.09	445.45	437.84
2011	379.82	412.23	429.82	438.33	416.40	433.05	437.43
2010	408.51	434.36	430.45	442.61	432.1	438.66	472.62

Table 5. Average of real and predicted values by the methods for Kermanshah in different years.

After implementing SVR, RF, and KNN using the Scikit-Learn library in Python, their predictions of chickpea yield were evaluated by computing RMSE, MBE, MAE, MAPE, and CC in the two defined scenarios (using an optimal combination of months and using the average of the growing period) with respect to the values reported by the Statistical Center and the Ministry of Agriculture of Iran. Tables 6 and 7 show the values of these validation measures in the (A) and (OC) scenarios. To compute these measures with the LOOCV technique, we generated eight sets of results for each year between 2010 and 2017. For each year, we first took the data of that year out of the dataset and trained the methods with the rest of the data. We then used the method to predict the removed data and compared the predictions with the actual data. This process was carried out for both (A) and (OC) scenarios. The results showed that among the implemented machine learning methods, RF with a CC of 0.86, RMSE of 40.71 and 43.39 (Kg/ha), and MAPE of 7–8% for the (A) and (OC) scenarios had the best accuracy in estimating the yield of rainfed chickpea farms. The least accurate method was KNN, which had a CC of 0.20 and 0.32, RMSE of 90.16 and 93.66 (Kg/ha), and MAPE of 17–19% in the two scenarios. SVR was found to have a CC of 0.46 and 0.52, RMSE of 73.57 and 80.94 (Kg/ha), and MAPE of 11–15% in the (A) and (OC) scenarios. Similar results were obtained in terms of MAE and MBE, showing RF to be the most accurate, SVR to be the second most accurate, and KNN to be the least accurate of the implemented methods. All methods performed better in all evaluation measures in the OC scenario, which indicates that this method (the optimal combination of months) is more appropriate for predicting the yield of rainfed chickpeas.

Table 6. Validity assessment of the implemented machine learning methods in the OC scenario.

Methods	MBE (Kg/ha)	RMSE (Kg/ha)	MAE (Kg/ha)	MAPE (%)	CC
SVR	48.92	73.57	49.71	11.94	0.46
RF	32.31	40.71	33.00	7.96	0.86
KNN	72.85	90.16	73.72	17.18	0.20

Table 7. Validity assessment of the implemented machine learning methods in the A scenario.

Methods	MBE (Kg/ha)	RMSE (Kg/ha)	MAE (Kg/ha)	MAPE (%)	CC
SVR	63.77	80.94	63.77	15.92	0.52
RF	34.88	43.39	34.89	8.32	0.86
KNN	80.44	93.66	80.44	19.56	0.32

Figure 6a–f show the difference between the actual values and the prediction made by the methods in the A and OC scenarios. For both scenarios, the line drawn for RF (Figure 6c,d) is closest to the 1:1 line, reflecting the better performance of this method in predicting the yield of rainfed chickpea farms. Given that the lowest chickpea yield has occurred in 2015 due to drought and the highest yield has occurred in 2014 because of heavy Mediterranean rainfall, KNN has performed poorly in predicting the yield of both of these years in both scenarios. This is evident in the lines plotted for KNN in Figure 6e,f. In Figure 6a,b, it can be seen that SVR has not performed as poorly as KNN, but has failed to produce acceptable predictions for the yield in 2014 (the year with the highest yield in the data). Overall, RF has had the most reliable performance in predicting chickpea yield even in the presence of yield fluctuations due to rainy years and droughts.



Figure 6. Results of the prediction models.

The use of machine learning techniques for predicting crop yields can present new opportunities for decision makers to increase food security. Especially in areas where crop yields are affected by uncertain conditions, like rainfed agriculture. In rainfed agriculture, due to the fact that the plant is not irrigated by the farmer and the plant receives its required moisture from the environment, and these environmental conditions are variable. Use of these influential factors and machine learning techniques for predicting crop yields provides useful information to government decision makers, warehousing and farmers.

Remote sensing data are comprehensive, accurate and varied. However, they are only available in developed countries. The needs of these countries determine the priority of the time and place of collection of this data. Also due to the fact that the collection of this information is done from different satellites and different sensors. There are different scales, units and time periods, which prolongs the process of refining and collecting data. Meteorological data obtained from synoptic stations have less lost information and are easier to access for less developed countries because they are collected locally. This data also needs to be refined. Access to them in the studied area was difficult. Also, some sensors were not present in all stations. Considering two different time scenarios for data collection as well as combining data for months shows the impact of some indicators is greater in certain seasons. For example, in the RF method for the OC time scenario, the most important criteria were surface moisture, minimum air humidity, and subsurface humidity in March, January, and January, respectively. For SVR and KNN methods, the most important indicators were ET and GPP, which were MA (average March and April) and April, respectively. The application of machine learning methods using the collected data showed that the random forest method had the least errors for the OC scenario. In summary, defining different time scenarios for data collection enables the researcher to make more accurate estimates of future rainfed agricultural yield by collecting specific data at specific times.

6. Conclusions

The rising world population and the consequent increase in demand for reliable food supplies and also the impact of climate change on food production have rekindled concerns about food security in many parts of the world. In many areas that lack sufficient water resources, rainfed farms are the main source of food and income, but these farms are being increasingly affected by climate change. Predicting the yields of these farms based on climate and remote sensing data can greatly guide government policies and help farmers and agriculture-related businesses make better decisions regarding resource procurement and allocation. In this study, the yield of rainfed chickpea farms in 11 top chickpea producing counties in Kermanshah province of Iran was predicted using SVR, RF, and KNN. The predictions were then evaluated in terms of multiple measures. Given the sensitivity of rainfed chickpea yield to the time of data, the predictions were made in two scenarios: (1) using the averages of the data of all growing months; and (2) using the data of a combination of months with the highest correlation with the yield. A total of 16 factors that tend to affect plant growth were collected from weather and remote sensing data, and an integrated database of these factors was formed. To improve prediction accuracy, the raster maps of rainfed farms were merged with slope and altitude maps of the province. Overall, RF showed the highest prediction accuracy in terms of all computed validation measures. The obtained predictions were only 7–8% different from the statistics reported by the Statistical Center and the Ministry of Agriculture of Iran, indicating that machine learning methods can indeed be a reliable choice for crop yield modeling. In future studies, it could be worthwhile to examine the effect of genotype and seed type of spring and autumn chickpea cultivated in local farms and also the soil type of these farms on the yield prediction at farm and regional level and also to explore the economic dimension of this discussion. Uncertainties of this study are related to amount of rainfall in the future which has been affected by global warming and environmental issues in the past two decades all over the world specially in Iran. We tried to predict production of chickpea in Kermanshah Province as one of the main producers of this product in Iran. Results of this research could help government policy makers to have an accurate statistic of chickpea. It is essential for import and export too.

Author Contributions: Conceptualization, S.R.; Methodology, S.R.; Software, E.J.; Validation, A.M., M.J.; Formal analysis, E.J.; Investigation, A.M.; Resources, A.I. and M.R.; Data curation, A.I. and M.J.; Writing—original draft preparation, S.R.; Writing—review and editing, S.C. and A.I.; Visualization,

M.R.; Supervision, A.M.; Project administration, K.T.; Funding acquisition, S.C. and K.T. All authors have read and agreed to the published version of the manuscript.

Funding: Not Applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This research was supported by the Prince of Songkla University from grant number-ENV6402012N.

Conflicts of Interest: The authors declare no conflict of interest.

Future Study: Researchers in the future can use different crops for prediction, use other techniques of prediction, and investigate other locations for chickpea and other products.

Nomenclature

BPNN	Back-Propagation Neural Networks
CART	Classification and Regression Trees
CCI	Climate Change Initiative
CHIRPS	Climate Hazards Group InfraRed Precipitation
CC	Correlation Coefficient
CHPclim	Climate Hazard group Precipitation climatology
DL	Deep Learning
DEM	Digital Elevation Model
EVI	Enhanced Vegetation Index
ESA	European Space Agency
ET	Evapotranspiration
ERT	Extremely Randomized Trees
FPAR	Fraction of Photosynthetically Active Radiation
GPR	Gaussian Process Regression
GFSAD	Global Food-Support Analysis Data
GPP	Gross Primary Production
LAI	Leaf Area Index
LOOCV	Leave One Out Cross Validation
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
MODIS	Moderate Resolution Imaging Spectroradiometer
MLR	Multiple Linear Regression
NASA	National Aeronautics and Space Administration
KNN	K-Nearest Neighbors
RF	Random Forest
RMSE	Root Mean Square Error
SM	Soil Moisture
SMAP	Soil Moisture Active Passive
SOM	Soli Organic Matter
SVM	Support Vector Machine
SVR	Support Vector Regression
NDVI	Normalized difference vegetation index
TIR	Thermal infrared
USDA	United States Department of Agriculture
USGS	United States Geological Survey

Appendix A

Tables A1–A8 are data for real and prediction values of dryland chickpea yield for average scenario from 2010 to 2017.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Tomoroda min (8C)	6.2	14.2	10.0	11.0	6.2	16.0	0.1	11 5	8.0	12.2	6.1
Temperature min (°C)	0.5	14.5	10.0	11.0	0.2	10.9	9.1	11.3	0.0	15.5	0.1
Temperature max (°C)	24.0	30.2	23.5	24.6	23.6	31.5	23.9	24.6	24.4	30.3	22.8
Average temperature (°C)	15.4	21.4	17.2	17.6	15.2	24.5	16.6	18.1	18.6	21.9	14.8
Precipitation (mm)	2.1	1.8	2.1	2.2	1.8	0.7	2.3	1.7	2.3	1.2	1.7
Humidity max (g/Kg)	76.8	59.9	59.8	61.7	81.1	57.1	68.1	58.4	69.7	68.7	71.9
Humidity min (g/Kg)	29.4	29.1	27.9	27.4	30.9	22.6	26.8	30.0	28.2	25.8	29.7
Average humidity (g/Kg)	52.6	44.2	42.5	44.3	54.1	38.7	45.9	44.1	44.2	47.0	49.8
Dry temperature (°C)	3.5	6.0	1.7	2.3	4.0	4.8	2.3	3.0	2.9	7.1	2.0
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.2	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	27.2	28.3	27.9	30.1	28.6	22.4	31.3	25.9	27.7	28.3	21.7
Leaf area index	5.9	5.9	6.1	6.5	6.6	4.3	7.2	5.4	6.2	5.8	4.6
Gross primary production (kg*C/m ²)	157.7	148.6	173.8	181.2	180.8	88.8	199.8	149.2	172.1	149.7	137.5
Evapotranspiration (kg/m^2)	88.2	75.9	92.5	99.9	94.0	36.9	106.3	94.1	94.3	73.4	87.2
Surface soil moisture (mm)	11.7	10.3	12.9	12.9	11.4	7.5	13.0	11.2	12.3	8.1	12.6
Subsurface soil moisture (mm)	57.0	46.0	57.3	56.9	50.2	28.4	62.8	49.7	64.9	33.6	55.7
Real value of production (Kg/ha)	622	519	534	411	550	525	696	508	617	590	400
Predicted production by RF (Kg/ha)	565	512	486	445	525	488	613	485	552	561	424
Predicted production by SVR (Kg/ha)	509	511	480	487	520	525	516	478	500	478	448
Predicted production by KNN (Kg/ha)	461	426	493	517	493	463	550	468	491	500	472

Table A1. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2017.

Table A2. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2016.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.9	13.8	9.5	10.8	5.9	17.2	9.1	11.9	7.7	14.1	5.8
Temperature max (°C)	23.4	28.9	22.9	23.8	23.2	31.1	23.2	24.2	24.0	30.0	22.2
Average temperature (°C)	15.5	22.0	18.2	17.2	14.8	24.4	16.1	18.0	18.2	21.9	14.4
Precipitation (mm)	1.7	2.0	2.2	2.3	2.2	1.7	2.5	2.0	2.2	1.3	1.8
Humidity max (g/Kg)	78.7	57.7	62.0	64.7	83.6	59.5	72.3	60.5	72.1	70.7	74.9
Humidity min (g/Kg)	30.1	30.0	29.4	30.1	31.4	23.2	29.2	31.2	28.6	26.9	30.8
Average humidity (g/Kg)	54.0	43.1	41.9	46.9	55.5	40.1	50.0	45.8	44.7	48.7	51.5
Dry temperature (° C)	4.2	6.7	2.8	3.6	4.1	6.2	3.6	4.1	3.3	8.1	2.7
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	34.2	33.3	34.3	39.0	35.4	25.9	39.7	30.7	35.7	33.4	28.7
Leaf area index	7.6	7.2	7.6	8.9	8.3	5.1	9.5	6.4	8.2	7.2	6.3
Gross primary production (kg*C/m ²)	230.8	222.6	239.1	266.9	245.1	168.1	284.4	206.2	248.9	223.4	191.1
Evapotranspiration (kg/m ²)	85.6	75.2	87.8	108.7	89.5	43.3	106.7	73.0	90.9	73.9	83.6
Surface soil moisture (mm)	12.7	10.1	13.9	12.7	10.6	7.6	13.7	9.7	13.4	8.3	12.4
Subsurface soil moisture (mm)	62.9	45.2	62.5	56.3	46.7	29.1	67.2	39.8	71.2	34.4	55.0
Real value of production (Kg/ha)	511	495	525	538	507	560	684	550	523	546	500
Predicted production by RF (Kg/ha)	523	503	525	557	517	469	631	528	542	533	487
Predicted production by SVR (Kg/ha)	511	512	509	538	507	487	566	514	523	518	500
Predicted production by KNN (Kg/ha)	505	543	515	531	468	558	546	563	531	543	445

Table A3. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2015.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.8	15.2	10.6	10.7	5.9	17.2	9.2	11.2	7.2	13.7	5.9
Temperature max (°C)	25.2	29.2	24.4	25.3	25.2	31.8	24.9	25.0	26.4	30.5	23.6
Average temperature (°C)	16.5	22.5	19.7	18.4	16.0	25.0	17.1	18.3	17.5	22.5	15.4
Precipitation (mm)	0.7	0.7	0.8	0.9	0.9	0.3	0.8	0.9	0.6	0.4	0.7
Humidity max (g/Kg)	69.5	48.4	48.5	53.7	73.3	48.0	60.2	53.0	55.9	60.2	62.3
Humidity min (g/Kg)	28.0	22.0	21.1	21.8	23.7	17.1	20.7	26.0	18.0	18.9	23.9
Average humidity (g/Kg)	47.1	34.1	31.4	36.1	45.5	30.1	38.7	38.8	35.1	38.3	40.6
Dry temperature (°C)	3.4	3.4	-0.8	0.0	2.0	2.3	-0.3	1.2	-1.4	4.3	-0.6
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.1	0.3	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	26.6	28.4	30.7	37.6	33.1	19.4	37.7	28.9	28.8	28.6	24.8
Leaf area index	5.3	5.8	6.6	8.6	7.5	3.5	9.0	6.0	6.2	5.9	5.1
Gross primary production (kg*C/m ²)	134.7	134.5	161.1	198.9	176.8	62.4	204.8	136.6	152.0	137.4	129.0
Evapotranspiration (kg/m ²)	46.8	50.3	59.7	83.3	63.9	19.4	79.5	55.6	54.7	52.1	50.1
Surface soil moisture (mm)	4.4	4.0	4.1	5.1	4.0	3.3	4.5	4.1	4.8	3.3	5.1
Subsurface soil moisture (mm)	21.7	18.1	18.4	23.7	17.6	13.0	22.4	17.0	26.5	14.0	23.7
Real value of production (Kg/ha)	300	270	306	250	290	230	394	235	305	225	295
Predicted production by RF (Kg/ha)	332	291	389	345	401	291	395	287	389	289	350
Predicted production by SVR (Kg/ha)	367	341	477	402	460	371	406	343	447	357	349
Predicted production by KNN (Kg/ha)	386	344	481	463	479	382	463	337	455	387	374

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.9	15.7	10.3	11.4	6.5	17.5	10.0	12.4	8.1	13.9	6.6
Temperature max (°C)	24.2	28.7	23.5	24.9	24.5	31.7	24.4	24.9	24.7	30.6	23.2
Average temperature (°C)	15.9	24.4	18.8	19.9	15.7	26.9	17.3	20.6	19.0	22.4	17.1
Precipitation (mm)	1.4	1.0	1.1	0.9	0.7	0.9	0.9	1.3	1.0	1.0	1.0
Humidity max (g/Kg)	72.4	47.8	57.3	54.5	78.9	49.9	62.6	51.3	67.4	65.9	68.1
Humidity min (g/Kg)	25.2	25.2	24.5	22.8	26.0	19.7	23.3	26.9	23.4	23.7	25.9
Average humidity (g/Kg)	48.1	33.9	37.3	35.4	50.5	30.5	41.7	36.5	39.9	44.4	41.6
Dry temperature (°C)	2.4	5.1	1.4	1.3	3.5	3.8	1.6	2.8	2.1	6.7	1.6
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.2	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	31.4	30.0	29.4	32.8	30.1	25.0	35.6	27.0	31.8	29.2	23.7
Leaf area index	6.8	6.3	6.2	6.9	6.8	5.0	8.1	5.3	7.0	6.0	4.9
Gross primary production (kg*C/m ²)	179.6	148.0	178.1	186.9	181.5	89.3	214.3	131.2	190.6	144.3	141.7
Evapotranspiration (kg/m^2)	78.1	63.6	75.9	82.9	76.4	37.7	88.4	58.2	80.3	60.2	70.4
Surface soil moisture (mm)	6.7	5.8	8.6	7.1	6.8	4.5	7.4	5.4	7.3	4.9	7.5
Subsurface soil moisture (mm)	32.2	28.4	38.4	34.3	31.3	17.8	39.0	22.4	39.0	20.8	35.7
Real value of production (Kg/ha)	608	554	661	477	522	500	685	530	552	762	405
Predicted production by RF (Kg/ha)	415	516	592	501	489	447	636	451	569	658	490
Predicted production by SVR (Kg/ha)	454	469	483	461	462	500	470	358	475	418	435
Predicted production by KNN (Kg/ha)	427	431	463	475	427	431	544	362	513	419	431

Table A4. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2014.

Table A5. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2013.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.5	13.5	9.8	10.6	6.0	17.1	9.4	11.6	8.0	13.8	6.0
Temperature max (°C)	23.7	27.8	22.8	23.8	23.9	30.9	23.6	23.8	24.5	29.7	22.8
Average temperature (°C)	15.6	23.4	18.3	18.8	15.1	26.1	16.4	19.3	18.6	21.9	16.3
Precipitation (mm)	0.5	0.5	0.8	1.4	0.6	0.3	0.8	0.9	0.7	0.6	0.8
Humidity max (g/Kg)	71.3	49.5	53.3	57.6	76.4	47.8	63.5	55.3	62.9	65.6	68.7
Humidity min (g/Kg)	23.7	24.8	23.7	25.6	25.8	17.1	24.8	26.8	21.9	21.5	26.3
Average humidity (g/Kg)	45.8	33.9	34.7	38.0	48.8	28.1	42.8	37.9	36.9	41.9	42.1
Dry temperature (° C)	1.5	4.4	0.2	1.8	2.5	2.9	1.6	2.2	1.0	5.7	0.9
Normalized difference vegetation index	0.3	0.2	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.2	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	27.2	26.0	29.3	36.3	32.1	20.3	36.5	25.7	30.1	26.3	25.0
Leaf area index	5.5	5.1	6.1	7.9	7.1	3.7	8.2	5.0	6.4	5.1	5.1
Gross primary production (kg*C/m ²)	156.6	135.0	170.4	212.5	183.3	75.8	221.2	133.5	175.4	137.5	142.2
Evapotranspiration (kg/m ²)	56.3	48.7	62.0	86.9	67.1	25.2	82.6	53.8	62.6	50.2	59.8
Surface soil moisture (mm)	5.4	5.1	7.1	7.4	5.6	5.1	7.9	5.1	6.1	4.9	6.9
Subsurface soil moisture (mm)	25.1	22.0	30.8	31.5	24.1	19.5	37.8	20.5	30.8	20.2	29.5
Real value of production (Kg/ha)	492	489	527	381	461	433	599	433	491	524	401
Predicted production by RF (Kg/ha)	486	429	502	404	451	399	587	416	483	454	419
Predicted production by SVR (Kg/ha)	449	346	509	437	461	417	489	341	491	375	403
Predicted production by KNN (Kg/ha)	478	362	499	484	455	398	543	366	472	402	459

Table A6. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2012.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.5	14.7	9.4	10.5	5.8	17.8	9.2	11.8	8.4	14.1	5.4
Temperature max (°C)	23.7	28.2	22.3	23.8	23.4	31.6	23.0	23.5	24.4	30.4	22.0
Average temperature (°C)	15.4	23.8	17.7	18.7	14.9	26.6	16.3	19.2	18.3	22.2	15.8
Precipitation (mm)	0.9	0.9	0.7	1.3	0.7	0.6	0.9	1.3	0.7	0.7	0.8
Humidity max (g/Kg)	69.1	44.8	50.3	53.8	74.4	42.5	58.5	46.1	65.0	65.0	68.6
Humidity min (g/Kg)	24.2	21.6	23.6	24.5	24.5	15.9	25.5	23.0	22.5	20.0	27.9
Average humidity (g/Kg)	45.3	29.9	34.2	35.9	47.0	25.2	36.0	32.6	34.0	33.0	43.1
Dry temperature (°C)	1.3	2.7	-0.5	0.7	1.7	1.3	0.6	-0.5	0.5	5.0	0.9
Normalized difference vegetation index	0.2	0.2	0.2	0.3	0.3	0.2	0.3	0.2	0.2	0.2	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	21.6	22.5	25.8	29.0	28.2	15.2	31.5	21.6	25.2	23.1	21.1
Leaf area index	4.2	4.4	5.3	6.2	6.2	2.6	7.2	4.1	5.3	4.5	4.4
Gross primary production (kg*C/m ²)	118.0	108.9	150.0	175.8	171.7	47.9	199.0	107.3	150.2	113.2	137.0
Evapotranspiration (kg/m ²)	48.7	43.1	61.9	79.5	70.4	14.6	83.6	47.8	61.9	46.6	64.3
Surface soil moisture (mm)	5.5	4.9	6.5	8.3	5.2	3.2	7.9	3.8	6.0	3.5	6.4
Subsurface soil moisture (mm)	24.3	20.3	27.6	35.9	21.7	12.2	36.7	15.2	29.8	14.1	26.7
Real value of production (Kg/ha)	487	484	521	377	456	428	593	428	485	518	397
Predicted production by RF (Kg/ha)	454	414	502	427	451	395	547	418	472	459	402
Predicted production by SVR (Kg/ha)	426	437	460	478	456	428	475	428	448	466	397
Predicted production by KNN (Kg/ha)	388	446	457	471	479	382	451	446	491	380	426

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	6.3	14.4	9.2	10.6	5.6	17.0	9.1	11.7	7.5	13.7	5.7
Temperature max (°C)	23.1	27.2	22.1	23.8	23.2	30.8	22.9	23.3	23.7	29.6	22.0
Average temperature (°C)	15.0	23.1	17.3	18.7	14.5	26.1	16.1	18.9	18.0	21.7	15.9
Precipitation (mm)	1.3	1.4	1.5	1.9	1.2	0.6	1.7	1.5	1.4	0.9	1.7
Humidity max (g/Kg)	79.0	50.3	61.3	57.8	81.0	47.3	66.0	51.9	67.8	66.0	74.4
Humidity min (g/Kg)	30.0	25.2	26.2	26.4	26.9	17.2	27.7	25.0	24.9	21.5	29.2
Average humidity (g/Kg)	52.6	34.7	40.6	38.8	52.9	27.8	45.6	36.0	40.9	43.0	46.0
Dry temperature (°C)	3.5	4.6	1.7	1.9	3.0	2.5	2.3	1.0	1.4	5.8	2.0
Normalized difference vegetation index	0.3	0.2	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.2	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	26.4	25.3	29.6	30.1	31.5	17.8	32.5	24.1	28.7	25.7	24.7
Leaf area index	5.5	5.1	6.4	6.5	7.1	3.2	7.5	4.7	6.3	5.0	5.4
Gross primary production (kg*C/m ²)	148.9	130.9	177.4	178.6	190.3	68.7	201.9	125.4	171.1	134.0	153.1
Evapotranspiration (kg/m^2)	68.7	54.9	76.0	83.6	81.5	21.9	88.7	54.7	75.4	52.7	75.9
Surface soil moisture (mm)	7.8	6.7	10.3	10.0	8.0	4.3	10.1	6.7	8.6	7.1	7.7
Subsurface soil moisture (mm)	35.2	27.9	42.9	41.2	31.9	15.0	46.6	25.6	42.0	28.3	31.4
Real value of production (Kg/ha)	374	372	401	490	351	329	456	329	373	398	305
Predicted production by RF (Kg/ha)	480	392	424	477	409	353	477	368	439	432	331
Predicted production by SVR (Kg/ha)	445	372	479	472	461	381	493	354	473	398	436
Predicted production by KNN (Kg/ha)	439	362	463	463	473	382	527	366	465	432	439

Table A7. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2011.

Table A8. Criteria values of each city and real and prediction values of dryland chickpea yield for average scenario 2010.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	7.0	15.1	10.6	11.5	7.1	18.5	10.0	12.7	8.3	14.7	7.2
Temperature max (°C)	24.7	28.1	23.4	24.6	24.3	32.5	24.1	24.3	24.9	30.2	22.9
Average temperature (°C)	16.1	23.9	18.8	19.6	15.7	27.7	16.9	20.0	18.9	22.4	17.1
Precipitation (mm)	1.4	2.3	1.8	2.0	1.7	1.4	2.0	2.0	1.5	2.1	1.4
Humidity max (g/Kg)	78.9	55.2	60.5	61.8	80.7	53.2	69.1	56.0	71.2	70.5	72.1
Humidity min (g/Kg)	26.4	29.6	27.3	30.3	29.9	20.8	28.3	30.4	26.7	28.0	29.2
Average humidity (g/Kg)	51.6	39.3	40.3	42.6	54.7	32.8	48.0	41.1	44.1	49.3	45.0
Dry temperature (°C)	4.1	7.2	2.7	4.1	4.8	5.6	3.6	3.8	3.3	8.5	2.6
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.3	0.3
Enhanced vegetation index	0.2	0.2	0.3	0.2	0.3	0.1	0.3	0.2	0.3	0.2	0.2
Fraction of photosynthetically active radiation	35.9	32.7	39.2	38.5	39.5	21.3	42.7	30.0	39.1	31.3	33.3
Leaf area index	8.1	7.1	9.2	9.0	9.7	3.9	11.0	6.2	9.5	6.5	7.6
Gross primary production (kg*C/m ²)	208.2	170.1	236.5	225.2	236.0	71.6	261.8	154.8	235.4	163.0	197.0
Evapotranspiration (kg/m ²)	87.0	73.9	97.4	105.5	95.2	29.8	113.8	71.7	98.0	70.9	95.8
Surface soil moisture (mm)	7.6	6.7	7.8	7.9	7.5	5.5	9.0	6.9	7.1	6.8	7.1
Subsurface soil moisture (mm)	34.1	27.4	31.8	31.8	30.9	19.3	40.9	25.6	33.7	26.6	29.1
Real value of production (Kg/ha)	423	420	453	327	396	372	515	372	421	450	345
Predicted production by RF (Kg/ha)	454	460	436	379	406	398	519	397	432	471	403
Predicted production by SVR (Kg/ha)	450	447	437	394	422	411	515	449	426	450	424
Predicted production by KNN (Kg/ha)	531	468	480	478	451	398	520	485	480	432	476

Appendix B

Tables A9–A16 are data for real and prediction values of dryland chickpea yield for optimum combination scenario from 2010 to 2017.

Table A9. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2017.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	11.1	21.5	15.9	17.9	10.1	24.0	15.1	17.8	12.9	19.4	10.6
Temperature max (°C)	27.5	32.5	27.1	28.3	26.9	35.8	27.6	28.2	27.7	34.7	26.1
Average temperature (°C)	6.7	12.6	8.7	8.1	6.9	14.7	7.8	9.0	9.1	13.2	6.2
Precipitation (mm)	5.5	4.4	4.7	5.5	4.0	1.6	6.0	4.3	4.8	3.2	3.2
Humidity max (g/Kg)	82.9	76.9	70.9	76.5	84.6	76.3	77.0	71.6	77.8	86.3	80.9
Humidity min (g/Kg)	14.8	13.4	11.9	10.7	16.6	7.6	11.1	11.9	10.9	12.0	12.4
Average humidity (g/Kg)	65.2	60.5	56.5	60.8	64.4	58.8	58.7	61.0	58.5	63.2	62.4
Dry temperature (°C)	4.6	5.9	2.5	2.8	5.0	2.7	3.4	3.0	4.1	6.8	3.6
Normalized difference vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1	0.2	0.2	0.1
Enhanced vegetation index	0.1	0.2	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.2	0.1
Fraction of photosynthetically active radiation	4.0	6.2	4.7	5.2	4.9	6.3	5.3	3.5	4.3	6.7	2.5
Leaf area index	21.1	30.1	23.3	25.7	23.4	31.3	25.6	19.0	21.9	32.3	13.6
Gross primary production (kg*C/m ²)	125.0	170.3	134.2	147.3	132.3	150.1	150.5	109.8	129.7	182.6	68.4
Evapotranspiration (kg/m ²)	97.6	105.6	108.4	127.4	104.4	60.6	124.1	111.7	106.7	112.3	100.3
Surface soil moisture (mm)	18.9	17.2	21.6	22.6	21.1	13.5	21.9	19.1	19.8	13.7	21.8
Subsurface soil moisture (mm)	9.4	8.9	8.9	8.7	8.6	8.2	10.2	8.6	11.4	8.5	9.4
Real value of production (Kg/ha)	622	519	534	411	550	525	696	400	508	617	590
Predicted production by RF (Kg/ha)	575	522	553	464	542	498	636	443	559	509	586
Predicted production by SVR (Kg/ha)	573	518	506	457	550	453	467	484	535	526	534
Predicted production by KNN (Kg/ha)	520	561	522	500	530	418	500	539	581	550	548

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	12.0	22.6	16.3	17.9	11.4	24.9	15.9	18.9	14.0	21.3	11.7
Temperature max (°C)	25.8	34.7	25.9	26.2	25.9	34.2	25.6	27.1	26.5	32.9	24.8
Average temperature (°C)	8.4	14.1	10.3	9.5	8.2	15.6	8.7	10.2	10.3	14.0	7.4
Precipitation (mm)	3.3	5.7	3.2	4.7	2.4	2.3	5.6	3.6	3.0	2.3	3.2
Humidity max (g/Kg)	85.9	68.9	73.7	74.7	90.5	74.3	80.6	67.6	82.7	83.6	85.3
Humidity min (g/Kg)	15.1	15.7	13.9	14.2	17.0	9.5	14.1	14.8	14.3	13.1	15.6
Average humidity (g/Kg)	63.9	56.2	52.4	58.5	63.2	57.5	60.0	58.3	55.0	62.5	60.2
Dry temperature (°C)	6.9	7.6	5.4	6.1	7.2	5.9	6.4	6.1	5.6	9.6	5.7
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2
Enhanced vegetation index	0.2	0.3	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2
Fraction of photosynthetically active radiation	7.5	9.0	7.8	8.8	8.3	7.8	9.7	8.2	8.5	9.4	5.1
Leaf area index	35.0	41.0	36.0	40.1	36.9	37.8	42.0	38.6	38.3	42.8	25.3
Gross primary production (kg*C/m ²)	353.2	382.1	354.4	393.7	362.4	322.4	421.3	353.1	379.4	385.9	244.4
Evapotranspiration (kg/m ²)	117.0	117.5	123.9	139.8	125.6	65.5	137.3	118.1	127.1	123.8	103.4
Surface soil moisture (mm)	20.2	16.0	21.8	20.5	16.1	12.2	21.9	16.0	20.6	12.9	19.7
Subsurface soil moisture (mm)	11.3	9.1	11.6	9.9	9.3	8.2	12.7	8.5	14.7	8.5	10.5
Real value of production (Kg/ha)	511	495	525	538	507	560	684	500	550	523	546
Predicted production by RF (Kg/ha)	525	513	575	537	511	460	635	503	557	521	540
Predicted production by SVR (Kg/ha)	510	495	511	538	507	465	570	500	545	523	546
Predicted production by KNN (Kg/ha)	514	519	514	549	514	514	549	514	521	521	511

Table A10. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2016.

Table A11. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2015.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature Min (°C)	13.9	21.8	18.1	18.4	11.6	24.4	17.4	18.2	13.3	20.6	11.2
Temperature max (°C)	29.0	33.2	27.9	28.8	28.5	36.0	28.4	29.2	29.7	34.7	27.0
Average temperature (°C)	7.8	13.7	9.9	9.1	7.6	15.8	8.2	9.4	8.6	13.8	6.6
Precipitation (mm)	2.0	2.5	2.4	2.9	2.3	1.2	2.4	2.9	1.6	1.6	1.9
Humidity max (g/Kg)	78.9	53.5	59.2	59.6	78.6	54.9	71.1	61.1	66.2	71.4	70.2
Humidity min (g/Kg)	20.2	12.9	9.7	9.1	13.7	8.1	9.6	11.6	7.3	10.2	10.1
Average humidity (g/Kg)	58.2	45.6	42.8	48.6	53.8	43.6	51.6	53.5	46.9	51.3	51.9
Dry temperature (°C)	3.9	3.4	0.2	0.8	3.6	1.2	0.7	1.6	-0.5	4.5	0.9
Normalized difference vegetation index	0.3	0.3	0.3	0.4	0.3	0.2	0.4	0.3	0.3	0.3	0.3
Enhanced vegetation index	0.2	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.2
Fraction of photosynthetically active radiation	6.6	8.0	8.2	11.0	9.3	4.8	12.2	8.3	8.0	8.2	5.9
Leaf area index	32.7	37.8	37.8	46.5	40.9	26.4	49.2	39.1	36.7	38.6	29.0
Gross primary production (kg*C/m ²)	177.8	199.9	198.5	243.9	212.8	112.5	255.9	203.4	194.3	205.0	133.8
Evapotranspiration (kg/m ²)	69.5	79.3	83.6	127.3	93.9	20.4	129.4	88.9	81.4	80.7	63.2
Surface soil moisture (mm)	6.4	5.5	5.9	7.7	5.8	4.1	6.7	5.7	6.9	4.1	7.7
Subsurface soil moisture (mm)	9.3	8.9	8.6	8.6	8.4	8.2	9.4	8.3	10.8	8.5	9.0
Real value of production (Kg/ha)	300	270	306	250	290	230	394	295	235	305	225
Predicted production by RF (Kg/ha)	336	310	282	301	370	285	375	332	299	345	289
Predicted production by SVR (Kg/ha)	437	331	333	378	414	346	394	329	362	351	377
Predicted production by KNN (Kg/ha)	445	386	386	412	448	350	418	398	375	375	358

Table A12. Criteria	a values of each cit	y and real and	prediction v	values of dry	land chickpea	yield for optim	um combinatio	m
scenario 2014.								

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	12.6	21.9	16.1	18.7	11.9	24.0	17.4	18.5	13.1	20.1	12.6
Temperature max (°C)	27.2	32.4	26.6	28.0	27.0	35.6	27.7	28.3	27.9	34.8	26.1
Average temperature (°C)	8.3	15.3	10.6	11.2	8.6	17.4	9.3	11.7	10.8	14.0	9.0
Precipitation (mm)	4.5	3.2	2.5	2.3	1.4	3.0	3.0	4.2	2.5	3.7	2.8
Humidity max (g/Kg)	84.7	57.5	70.7	69.1	85.3	62.7	75.8	64.6	79.6	79.6	76.8
Humidity min (g/Kg)	12.7	13.3	11.3	10.9	16.5	7.9	12.0	13.1	10.9	10.9	14.3
Average humidity (g/Kg)	61.1	46.6	48.8	47.5	56.6	45.5	53.7	49.7	51.6	59.9	51.3
Dry temperature (° C)	3.8	4.9	3.0	2.5	6.2	2.3	3.3	3.7	3.3	7.2	3.0
Normalized difference vegetation index	0.3	0.3	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	7.0	8.0	6.3	7.4	6.7	7.6	9.1	6.9	7.5	8.1	4.3
Leaf area index	33.6	37.6	30.8	35.9	31.0	36.4	40.5	34.2	35.0	38.2	22.2
Gross primary production (kg*C/m ²)	192.0	194.7	174.6	196.7	170.6	154.7	228.3	175.4	200.8	198.3	116.0
Evapotranspiration (kg/m ²)	106.8	94.0	99.2	118.7	95.0	47.4	128.6	90.9	114.0	93.9	88.7
Surface soil moisture (mm)	10.8	10.4	12.8	11.8	10.5	7.1	12.9	8.7	11.5	7.8	11.4
Subsurface soil moisture (mm)	11.4	9.5	17.9	12.7	14.1	8.2	13.3	8.8	15.2	8.5	17.9
Real value of production (Kg/ha)	608	554	661	477	522	500	685	405	530	552	762
Predicted production by RF (Kg/ha)	514	520	621	489	503	438	635	399	550	430	688
Predicted production by SVR (Kg/ha)	529	384	511	477	522	500	456	452	522	432	569
Predicted production by KNN (Kg/ha)	517	361	517	485	501	399	416	444	500	416	529

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	12.7	20.4	16.2	17.8	11.4	24.1	16.5	18.2	13.6	20.6	10.4
Temperature max (°C)	23.7	28.2	23.2	23.9	24.2	31.8	23.9	25.1	24.8	30.4	23.5
Average temperature (°C)	9.0	16.3	11.5	12.0	9.1	18.5	9.8	12.2	11.6	14.8	10.0
Precipitation (mm)	0.0	0.1	0.4	0.3	0.3	0.0	0.2	0.5	0.3	0.1	1.0
Humidity max (g/Kg)	78.0	50.3	58.5	63.8	84.4	50.4	70.0	60.9	69.1	74.0	75.3
Humidity min (g/Kg)	12.9	12.8	10.6	11.2	14.3	7.6	10.7	12.5	10.6	10.4	13.5
Average humidity (g/Kg)	49.2	37.8	39.4	42.3	52.0	32.9	47.1	43.9	40.7	48.6	47.6
Dry temperature (°C)	6.1	8.5	4.4	6.8	6.2	6.9	6.5	6.4	5.3	9.8	4.3
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.2
Enhanced vegetation index	0.2	0.2	0.2	0.2	0.3	0.2	0.3	0.2	0.2	0.2	0.2
Fraction of photosynthetically active radiation	6.2	6.8	6.9	8.4	8.5	5.3	8.9	6.4	7.5	6.7	4.9
Leaf area index	30.9	33.9	33.5	39.5	38.5	28.4	40.6	32.7	35.4	34.2	25.1
Gross primary production (kg*C/m ²)	172.6	178.4	175.9	213.2	191.8	122.1	216.3	172.9	189.4	181.2	115.2
Evapotranspiration (kg/m ²)	72.0	66.8	72.5	106.7	83.9	25.7	98.4	72.7	79.9	66.3	60.8
Surface soil moisture (mm)	6.3	6.5	9.3	10.2	7.3	6.2	10.0	6.9	7.8	6.8	9.3
Subsurface soil moisture (mm)	13.5	11.9	16.9	14.8	16.2	9.1	18.7	10.4	16.9	9.7	14.5
Real value of production (Kg/ha)	492	489	527	381	461	433	599	401	433	491	524
Predicted production by RF (Kg/ha)	481	443	486	424	450	410	545	401	439	451	466
Predicted production by SVR (Kg/ha)	470	458	469	416	461	405	432	456	433	477	436
Predicted production by KNN (Kg/ha)	441	393	430	450	407	350	463	441	425	441	433

Table A13. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2013.

Table A14. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2012.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	14.0	22.8	17.3	19.1	12.9	26.1	19.3	19.9	16.6	22.2	12.0
Temperature max (°C)	28.3	32.6	26.4	27.8	27.3	35.8	27.1	27.7	28.7	34.5	26.1
Average temperature (°C)	4.4	11.9	6.5	6.8	4.9	14.5	4.7	7.3	6.9	10.9	4.7
Precipitation (mm)	2.2	2.8	1.3	3.1	1.1	1.6	1.5	3.5	0.9	2.3	1.7
Humidity max (g/Kg)	79.9	56.4	65.4	66.7	87.7	50.9	77.9	55.7	79.3	77.8	83.6
Humidity min (g/Kg)	11.7	10.9	12.4	10.3	14.8	6.7	12.2	9.5	8.6	9.6	13.2
Average Humidity (g/Kg)	57.6	40.7	45.9	48.1	55.7	36.2	53.4	44.8	49.8	53.6	55.2
Dry temperature (°C)	3.6	4.1	0.3	3.0	3.6	1.4	2.7	1.2	2.3	6.8	2.8
Normalized difference vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.1
Enhanced vegetation index	0.1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
Fraction of photosynthetically active radiation	3.5	4.9	4.8	5.8	5.3	3.4	7.0	4.7	5.0	5.3	3.1
Leaf area index	19.0	25.1	24.2	28.0	25.3	19.5	31.7	24.8	24.5	27.3	15.9
Gross primary production (kg*C/m ²)	115.6	139.5	142.1	170.7	146.6	81.6	194.2	134.3	147.9	148.6	94.5
Evapotranspiration (kg/m ²)	61.3	67.0	77.5	107.7	87.2	20.5	115.0	71.1	82.1	73.1	75.9
Surface soil moisture (mm)	7.8	8.4	9.4	15.2	5.3	4.5	14.4	5.5	9.4	4.6	7.6
Subsurface soil moisture (mm)	10.3	10.0	11.6	10.9	11.5	8.9	13.7	9.1	13.1	8.9	13.4
Real value of production (Kg/ha)	487	484	521	377	456	428	593	397	428	485	518
Predicted production by RF (Kg/ha)	457	408	482	420	456	386	566	379	436	379	499
Predicted production by SVR (Kg/ha)	484	393	412	480	456	428	532	397	428	411	518
Predicted production by KNN (Kg/ha)	392	366	422	525	403	365	526	385	435	442	497

Table A15. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2011.

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	12.3	22.9	16.6	18.6	11.3	25.0	16.8	19.2	14.0	21.1	10.8
Temperature max (°C)	24.6	29.7	24.2	25.7	25.0	33.6	24.5	25.0	25.7	32.2	23.5
Average temperature (°C)	6.9	14.0	8.5	10.3	6.4	17.0	8.2	10.3	9.2	13.0	7.3
Precipitation (mm)	1.6	1.0	1.9	1.5	1.9	0.7	1.3	0.7	1.9	0.9	1.5
Humidity max (g/Kg)	85.0	58.4	67.6	67.0	83.3	56.8	74.9	60.8	77.3	77.8	74.7
Humidity min (g/Kg)	16.4	11.9	16.0	11.2	14.0	7.2	12.2	10.6	8.6	9.6	13.2
Average humidity (g/Kg)	59.0	42.9	48.2	46.4	58.6	37.1	53.4	43.2	49.8	53.6	52.3
Dry temperature (°C)	7.4	7.7	4.3	6.1	6.6	4.1	6.3	5.2	5.5	8.9	6.0
Normalized difference vegetation index	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.1
Enhanced vegetation index	0.1	0.2	0.2	0.1	0.2	0.2	0.2	0.2	0.1	0.2	0.1
Fraction of photosynthetically active radiation	3.9	4.9	5.5	4.5	6.0	4.2	5.4	4.5	4.8	5.4	2.9
Leaf area index	21.1	25.5	27.3	23.8	28.8	23.3	26.5	24.3	24.3	28.1	15.9
Gross primary production (kg*C/m ²)	116.2	133.0	145.7	121.2	153.3	106.6	138.6	125.6	129.2	145.1	72.1
Evapotranspiration (kg/m ²)	74.2	65.7	84.2	81.6	84.3	27.7	82.4	64.6	77.4	65.5	72.2
Surface soil moisture (mm)	10.4	9.0	15.9	15.4	10.2	5.8	15.1	8.8	12.0	10.2	9.1
Subsurface soil moisture (mm)	11.1	9.8	13.9	11.0	11.1	8.2	12.9	8.8	13.2	8.8	12.0
Real value of production (Kg/ha)	374	372	401	490	351	329	456	305	329	373	398
Predicted production by RF (Kg/ha)	445	394	446	471	398	356	475	331	417	390	412
Predicted production by SVR (Kg/ha)	541	372	438	489	445	342	451	372	457	400	420
Predicted production by KNN (Kg/ha)	535	385	414	477	436	350	414	385	464	421	541

Criteria	Eslamabad-e- Gharb	Gilanegharb	Harsin	Javanrud	Kangavar	Qasreshirin	Ravansar	Salas	Kermanshah	Sarpol-e- Zahab	Sonqor
Temperature min (°C)	12.2	22.6	17.0	18.5	11.6	26.5	16.7	19.7	13.4	21.8	11.9
Temperature max (°C)	26.2	29.9	24.5	25.9	25.5	35.0	25.3	25.7	26.1	32.5	23.9
Average temperature (°C)	10.0	16.8	12.5	12.9	10.4	19.7	11.0	13.3	12.7	15.2	11.4
Precipitation (mm)	2.3	2.8	2.1	2.8	1.2	2.1	2.3	3.1	2.0	2.4	1.5
Humidity max (g/Kg)	86.4	68.4	75.2	80.8	88.1	75.7	86.3	72.8	88.0	87.6	83.7
Humidity min (g/Kg)	14.2	15.4	12.6	11.4	15.6	7.1	11.0	11.4	9.5	11.7	12.7
Average humidity (g/Kg)	59.4	50.4	49.1	54.7	61.8	47.1	59.9	53.8	54.7	65.1	54.6
Dry temperature (°C)	6.4	8.2	4.3	5.9	6.4	5.2	5.4	5.1	5.7	8.9	3.7
Normalized difference vegetation index	0.3	0.3	0.3	0.3	0.3	0.2	0.4	0.3	0.4	0.3	0.3
Enhanced vegetation index	0.3	0.3	0.3	0.2	0.3	0.2	0.3	0.2	0.3	0.2	0.2
Fraction of photosynthetically active radiation	9.4	9.0	11.3	10.0	11.3	5.5	13.5	7.8	11.9	8.3	7.5
Leaf area index	41.2	40.5	47.0	43.2	46.1	28.9	51.1	37.3	47.7	39.3	34.6
Gross primary production (kg*C/m ²)	245.9	224.7	269.5	241.9	253.1	118.6	294.4	200.3	279.0	217.0	171.4
Evapotranspiration (kg/m ²)	133.2	114.1	150.9	155.6	138.5	35.0	173.5	113.7	156.3	107.4	132.0
Surface soil moisture (mm)	10.4	8.8	9.4	10.5	8.4	6.7	12.4	9.3	9.0	8.8	7.7
Subsurface soil moisture (mm)	12.5	9.4	19.6	11.2	20.3	8.2	17.7	8.6	16.9	8.6	17.0
Real value of production (Kg/ha)	423	420	453	327	396	372	515	345	372	421	450
Predicted production by RF (Kg/ha)	472	412	472	407	422	374	530	422	393	416	459
Predicted production by SVR (Kg/ha)	416	420	443	369	396	372	515	472	441	441	450
Predicted production by KNN (Kg/ha)	418	459	409	418	438	359	409	500	409	541	507

Table A16. Criteria values of each city and real and prediction values of dryland chickpea yield for optimum combination scenario 2010.

References

- Roselaar, S. Agriculture in Republican Italy. In *A Companion to Ancient Agriculture*; Wiley: New York, NY, USA, 2020; pp. 417–430.
 Amaratunga, V.; Wickramasinghe, L.; Perera, A.; Jayasinghe, J.; Rathnayake, U. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Math. Probl. Eng.* 2020, 2020, 1–11. [CrossRef]
- Bozorg-Haddad, O.; Sohrabi, S.; Delpasand, M.; Loáiciga, H.A. Dryland farming improvement by considering the rela-tion between rainfall variability and crop yield. *Environ. Dev. Sustain.* 2020, 23, 5316–5327. [CrossRef]
- 4. FAO. Available online: http://www.fao.org/americas/noticias/ver/en/c/409536 (accessed on 1 January 2021).
- 5. Kumar, J.; Abbo, S. Genetics of flowering time in chickpea and its bearing on productivity in semiarid environments. *Adv. Agron.* **2001**, *72*, 107–138.
- 6. Merga, B.; Haji, J. Economic importance of chickpea: Production, value, and world trade. *Cogent Food Agric.* **2019**, *5*, 1615718. [CrossRef]
- 7. ANRIS. Available online: http://anris.agri-peri.ir (accessed on 2 January 2021).
- 8. Alijani, B.; Harman, J.R. Synoptic Climatology of Precipitation in Iran. Ann. Assoc. Am. Geogr. 1985, 75, 404–416. [CrossRef]
- 9. Taheri, K.; Taheri, M.; Parise, M. Impact of intensive groundwater exploitation on an unprotected covered karst aquifer: A case study in Kermanshah Province, western Iran. *Environ. Earth Sci.* **2016**, *75*, 1221. [CrossRef]
- Gandhi, N.; Armstrong, L.J.; Petkar, O.; Tripathy, A.K. Rice crop yield prediction in india using support vector ma-chines. In Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 13–15 July 2016; pp. 1–5.
- 11. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.; Ciampitti, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* **2020**, *284*, 107886. [CrossRef]
- 12. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, 274, 144–159. [CrossRef]
- 13. Ejaz, N.; Abbasi, S. Wheat yield prediction using neural network and integrated svm-nn with regression. *Pak. J. Eng. Technol. Sci.* **2020**, *8*, 77–97.
- 14. Medar, R.A.; Rajpurohit, V.S.; Ambekar, A.M. Sugarcane Crop Yield Forecasting Model Using Supervised Machine Learning. *Int. J. Intell. Syst. Appl.* **2019**, *11*, 11–20. [CrossRef]
- 15. Chen, H.; Wu, W.; Liu, H.-B. Assessing the relative importance of climate variables to rice yield variation using support vector machines. *Theor. Appl. Clim.* **2016**, 126, 105–111. [CrossRef]
- 16. Kouadio, L.; Deo, R.C.; Byrareddy, V.; Adamowski, J.F.; Mushtaq, S.; Nguyen, V.P. Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. *Comput. Electron. Agric.* **2018**, *155*, 324–338. [CrossRef]
- 17. Kim, N.; Lee, Y.-W. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2016**, *34*, 383–390. [CrossRef]
- 18. Narasimhamurthy, V.; Kumar, P. Rice crop yield forecasting using random forest algorithm. *Int. J. Res. Appl. Sci. Eng. Technol. IJRASET* 2017, *5*, 1220–1225. [CrossRef]
- 19. Mohammadi, K.; Mostafaeipour, A.; Dinpashoh, Y.; Pouya, N. Electricity generation and energy cost estimation of large-scale wind turbines in Jarandagh, Iran. J. Energy 2014, 37. [CrossRef]

- 20. Zarezade, M.; Mostafaeipour, A. Identifying the effective factors on implementing the solar dryers for Yazd province, Iran. *Renew. Sustain. Energy Rev.* **2016**, *57*, 765–775. [CrossRef]
- 21. Rezaei, M.; Salimi, M.; Momeni, M.; Mostafaeipour, M. Investigation of the socio-economic feasibility of installing wind turbines to produce hydrogen: Case study. *Int. J. Hydrogen Energy* **2018**, *43*, 23135–23147. [CrossRef]
- 22. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* 2020, *12*, 236. [CrossRef]
- 23. Parviz, L. Assessing accuracy of barley yield forecasting with integration of climate variables and support vector regres-sion. *Ann. Univ. Mariae Curie-Sklodowska Sect. C Biol.* **2019**, *73*, 19–30. [CrossRef]
- 24. Kuwata, K.; Shibasaki, R. Estimating crop yields with deep learning and remotely sensed data. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 858–861.
- 25. Tiwari, P.; Shukla, P. Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors. In *Advances in Human Factors, Business Management, Training and Education*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 585–594.
- Sharifi, A. Yield prediction with machine learning algorithms and satellite images. *J. Sci. Food Agric.* 2021, *101*, 891–896. [CrossRef]
 Mostafaeipour, A.; Jooyandeh, E. Prioritizing the locations for hydrogen production using a hybrid wind-solar system: A case study. *Adv. Energy Res.* 2017, *5*, 107.
- 28. Li, B.; Tang, H.; Chen, D. Drought Monitoring Using the Modified Temperature/Vegetation Dryness Index. In Proceedings of the 2009 2nd International Congress on Image and Signal Processing, Tianjin, China, 17–19 October 2009; pp. 1–5. [CrossRef]
- 29. Pekin, B.; Macfarlane, C. Measurement of crown cover and leaf area index using digital cover photography and its ap-plication to remote sensing. *Remote Sens.* 2009, *1*, 1298–1320. [CrossRef]
- 30. Ritchie, J.T. Efficient Water Use in Crop Production: Discussion on the Generality of Relations Between Biomass Production and Evapotranspiration. In *Limitations to Efficient Water Use in Crop Production;* Wiley: New York, NY, USA, 2015; pp. 29–44.
- 31. Trejo, F.J.P.; Barbosa, H.A.; Peñaloza-Murillo, M.A.; Moreno, M.A.; Farías, A. Intercomparison of improved satellite rainfall estimation with CHIRPS gridded product and rain gauge data over Venezuela. *Atmósfera* **2016**, *29*, 323–342. [CrossRef]
- Saeidizand, R.; Sabetghadam, S.; Tarnavsky, E.; Pierleoni, A. Evaluation of CHIRPS rainfall estimates over Iran. Q. J. R. Meteorol. Soc. 2018, 144, 282–291. [CrossRef]
- 33. Vapnik, V.N. An overview of statistical learning theory. IEEE Trans. Neural Netw. 1999, 10, 988–999. [CrossRef] [PubMed]
- 34. Yousefi, M.; Khoshnevisan, B.; Shamshirband, S.; Motamedi, S.; Nasir, M.H.N.M.; Arif, M.; Ahmad, R. Support vector regression methodology for prediction of output energy in rice production. *Stoch. Environ. Res. Risk Assess.* 2015, 29, 2115–2126. [CrossRef]
- Cristianini, N.; Shawe-Taylor, J. An Introduction to Support vector Machines and Other Kernel-Based Learning Methods; Cambridge University Press: Cambridge, UK, 2000.
- 36. Dibike, Y.B.; Velickov, S.; Solomatine, D.; Abbott, M.B. Model Induction with Support Vector Machines: Introduction and Applications. J. Comput. Civ. Eng. 2001, 15, 208–216. [CrossRef]
- 37. Samui, P. Support vector machine applied to settlement of shallow foundations on cohesionless soils. *Comput. Geotech.* 2008, 35, 419–427. [CrossRef]
- 38. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. Classification and Regression Trees; CRC Press: Boca Raton, FL, USA, 1984.
- 39. Alpaydin, E. Introduction to Machine Learning; MIT Press: Cambridge, MA, USA, 2020.
- 40. Breiman, L. Bagging predictors. Mach. Learn. 1996, 24, 123-140. [CrossRef]
- Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 2015, 71, 804–818. [CrossRef]
- 42. Ghasemi, J.B.; Tavakoli, H. Application of random forest regression to spectral multivariate calibration. *Anal. Methods* **2013**, *5*, 1863–1871. [CrossRef]
- 43. Liaw, A.; Wiener, M. Classification and regression by random-forest. R News 2002, 2, 18–22.
- 44. Cover, T.; Hart, P. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 1967, 13, 21–27. [CrossRef]
- 45. Lee, K.-C. Radar target recognition by machine learning of k-nearest neighbors regression on angular diversity RCS. *Appl. Comput. Electromagn. Soc. J.* **2019**, *34*, 75–81.
- 46. Peterson, L.E. K-nearest neighbor. Scholarpedia 2009, 4, 1883. [CrossRef]
- 47. Mostafaeipour, A.; Goli, A.; Qolipour, M. Prediction of air travel demand using a hybrid artificial neural network (ANN) with Bat and Firefly algorithms: A case study. *J. Supercomput.* **2018**, *74*, 5461–5484. [CrossRef]
- 48. Mostafaeipour, A.; Fakhrzad, M.; Gharaat, S.; Jahangiri, M.; Dhanraj, J.; Band, S.; Issakhov, A.; Mosavi, A. Machine Learning for Prediction of Energy in Wheat Production. *Agriculture* **2020**, *10*, 517. [CrossRef]
- 49. Samadianfard, S.; Hashemi, S.; Kargar, K.; Izadyar, M.; Mostafaeipour, A.; Mosavi, A.; Nabipour, N.; Shamshirband, S. Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. *Energy Rep.* **2020**, *6*, 1147–1159. [CrossRef]
- Mostafaeipour, A.; Qolipour, M.; Goudarzi, H.; Jahangiri, M.; Golmohammadi, A.; Rezaei, M.; Goli, A.; Sadeghikhorami, L.; Sedeh, A.S.; Soltani, S.R.K. Implementation of adaptive neuro-fuzzy inference system (ANFIS) for performance prediction of fuel cell parameters. *J. Renew. Energy Environ.* 2019, *6*, 7–15.