Article

Soil Salinity Estimation for South Kazakhstan Based on SAR Sentinel-1 and Landsat-8,9 OLI Data with Machine Learning Models

Ravil I. Mukhamediev [1,2], Timur Merembayev [2], Yan Kuchin [1,2], Dmitry Malakhov [2,3], Elena Zaitseva [4], Vitaly Levashenko [4], Yelena Popova [5], Adilkhan Symagulov [1,2,*], Gulshat Sagatdinova [2] and Yedilkhan Amirgaliyev [2]

[1] Institute of Automation and Information Technology, Satbayev University (KazNRTU), Satpayev Str., 22A, Almaty 050013, Kazakhstan; r.mukhamediev@satbayev.university (R.I.M.); ykuchin@mail.ru (Y.K.)
[2] Institute of Information and Computational Technologies, Pushkin Str., 125, Almaty 050010, Kazakhstan; t.merembayev@iitu.edu.kz (T.M.); dmitry.malakhov@zool.kz (D.M.); alchemdzz@gmail.com (G.S.); amired@iict.kz (Y.A.)
[3] Institute of Zoology SC MES RK, al-Faraby Av., 93, Almaty 050060, Kazakhstan
[4] Faculty of Management Science and Informatics, University of Zilina, Univerzitná 8215/1, 010 26 Žilina, Slovakia; elena.zaitseva@fri.uniza.sk (E.Z.); vitaly@kifri.fri.uniza.sk (V.L.)
[5] Baltic International Academy, 1/4 Lomonosov Str., LV-1003 Riga, Latvia; yelenagp@gmail.com
* Correspondence: a.symagulov@satbayev.university; Tel.: +7-707-819-47-00

Abstract: Climate change, uneven distribution of water resources and anthropogenic impact have led to salinization and land degradation in the southern regions of Kazakhstan. Identification of saline lands and their mapping is a laborious process associated with a complex of ground measurements. Data from remote sensing are widely used to solve this problem. In this paper, the problem of assessing the salinity of the lands of the South Kazakhstan region using remote sensing data is considered. The aim of the study is to analyze the applicability of machine learning methods to assess the salinity of agricultural lands in southern Kazakhstan based on remote sensing. The authors present a salinity dataset obtained from field studies and containing more than 200 laboratory measurements of soil salinity. Moreover, the authors describe the results of applying several regression reconstruction algorithms (XGBoost, LightGBM, random forest, Support vector machines, Elastic net, etc.), where synthetic aperture radar (SAR) data from the Sentinel-1 satellite and optical data in the form of spectral salinity indices are used as input data. The obtained results show that, in general, these input data can be used to estimate salinity of the wetted arable land. XGBoost regressor ($R^2 = 0.282$) showed the best results. Supplementing the radar data with the values of salinity spectral index improves the result significantly ($R^2 = 0.356$). For the local datasets, the best result shown by the model is $R^2 = 0.473$ (SAR) and $R^2 = 0.654$ (SAR with spectral indexes), respectively. The study also revealed a number of problems that justify the need for a broader range of ground surveys and consideration of multi-year factors affecting soil salinity. Key results of the article: (i) a set of salinity data for different geographical zones of southern Kazakhstan is presented for the first time; (ii) a method is proposed for determining soil salinity on the basis of synthetic aperture radar supplemented with optical data, and this resulted in the improved prediction of the results for the region under consideration; (iii) a comparison of several types of machine learning models was made and it was found that boosted models give, on average, the best prediction result; (iv) a method for optimizing the number of model input parameters using explainable machine learning is proposed; (v) it is shown that the results obtained in this work are in better agreement with ground-based measurements of electrical conductivity than the results of the previously proposed global model.

Keywords: soil salinity; synthetic aperture radar; machine learning; regression; XGBoost; LightGBM; random forest; support vector machines; elastic net

1. Introduction

Climate change and anthropogenic impact lead to the degradation of agricultural land due to increased salinity. Climatic factors create a potential danger for soil salinization, while population growth and uncontrolled intensification of the use of land and water resources can lead to an abrupt reduction of pastures and arable land's area [1]. According to [2], the total area of primary saline soils is about 955 million hectares, and about 77 million hectares are exposed to secondary salinization, of which 58% are irrigated areas. Nearly 20% of all irrigated land is saline, and this proportion tends to increase despite significant land reclamation efforts. It is assumed that, by 2050, the salinity of arable land in the world may exceed 50% [3].

Irrigation of the salinized arable lands and degradation of agricultural lands in the South of Kazakhstan is a systemic negative factor that most strongly affects four regions of Kazakhstan: Turkestan, Almaty, Zhambyl and Kyzylorda. The main problem is related to water resources, which are formed by the transboundary flow of large rivers in the region (the Syrdarya, Ile and Chu rivers). Kazakhstan is located in the lower reaches of river basins and is very vulnerable accordingly with regards to a water supply. The growth of water consumption in the upper parts of the river basins belonging to the territories of neighboring countries (Uzbekistan, China, Kyrgyzstan) and climate change create problems in the water supply of irrigated arable land in southern Kazakhstan. The problem of salinization and its mapping appeared due to the low ensuring of water and food security in the south of Kazakhstan, where two of the big cities of the Republic are located (Almaty, Shymkent). There are several modern solutions that are based on remote sensing monitoring and GIS and have provided solutions for soil salinity.

One such direction is the application of synthetic aperture radar data for estimation of the salinity of various territories [4,5], as well as combined methods, including both radar and conditionally optical remote sensing data [6–8].

In this paper, the authors used the data from the field studies sampled during May–July 2022 in three geographically significantly separated areas of southern Kazakhstan. The synthetic aperture radar (SAR) data from Sentinel-1 satellite and optical data from Landsat-8 in the form of spectral salinity indices were used as remote sensing data.

The aim of the study was to analyze the applicability of these data for estimating the salinity of the territories, using machine learning methods. At the same time, the applicability of this method was evaluated to territories that differ significantly in geography and soil condition.

The second task was a comparative analysis of machine-learning algorithms, which could be used in the future to calculate the salinity of the territories of Kazakhstan.

The main contribution of this study is as follows:

1. A set of field data for the study of salinity in the southern regions of Kazakhstan was prepared;
2. The method of soil salinity assessment based on SAR data proposed earlier in the literature was supplemented;
3. The method of salinity assessment was extended by using a combination of multispectral and SAR radar data;
4. A comparative analysis of machine learning algorithms solving the salinity estimation problem on the proposed data set is performed;
5. The significant input parameters (features) were selected and their effect was estimated using the methods of explainable machine learning (EML);
6. The boundaries of the joint use of obtained field data were determined;
7. The obtained modeling results were compared with the results of one of the known models of soil salinity assessment.

Section 2 presents the discussion of scientific papers devoted to solving the salinity estimation problems using machine learning methods. Section 3 describes the salinity estimation methods based on several machine learning models. Section 4 presents the result. Interpretations of the findings are discussed in Section 5. The Section 6 summarizes the results, describes the limitations of the proposed method and formulates the objectives of further research.

2. Related Works

2.1. Classical Salinity Estimation Methods Based on Spectral Data

The use of remote sensing data to assess soil salinity has a history of more than 40 years [1]. After reviewing the scientific studies, the authors revealed that studies [9–11] demonstrate a similar approach to the issue of estimating the salinity level of arable irrigated areas; this estimation is based on satellite data and GIS technologies, including a description of the salinity of some areas of irrigated arable land in Kazakhstan [12,13]. Currently, the use of satellite data and GIS technologies is the most widely used means to monitor arable land salinity, which in some cases can compete successfully with technically complex and expensive ground-based monitoring methods. The ideas of various satellite monitoring techniques are generally similar [14–16]. Some satellite images, or their time series, are used from various orbital platforms [17–19], including hyperspectral data [20,21], low spatial resolution satellite data (such as MODIS) [22,23], and radar images [24]. The spectral channels of the visible and near infrared spectral ranges are used and analyzed for the detection of soil salinity [14–16]. The different indices of salinity are constructed in combination with the vegetation indices on the basis of channel combinations; this allows the creation of a logical system of relationships between the field samples data on the existing level of arable land salinity and the spectral characteristics of the underlying surface. Another approach is to use more complex methods, such as clustering algorithms for remote sensing data and other machine learning methods [25–27].

2.2. Machine Learning Methods in Salinity Estimation Problems

Machine learning methods (ML) are an important subsection of artificial intelligence (AI), putting into practice the ideas of AI to create learning systems [28]. Machine learning technologies can effectively solve the nonlinear problem of the relationship between soil properties and environmental factors [29]. There have been a number of serious results in solving problems of classification, clustering, and regression analysis in various scientific fields [30–37], and this is the consequence of using ML methods in the studies devoted to the soil salinity. starting around 2012.

In the paper [27], the authors provided an estimation and mapping of soil salinity in the Aral Sea basin; they used data such as terrain indices, height above sea level, remote sensing data, distance to drainage channels, long-term observations of groundwater levels, etc. These input variables predict salinity, and an electromagnetic induction instrument (CM-138) measures the electrical conductivity (EC) of the soil profile down to a depth of 1.5 m in vertical mode. The created classification model obtained an accuracy of estimating saline soils at close to 90%, the salinity threshold defined as about 0.7 dS m^{-1}. The authors provided a sensitivity analysis of the model; as a result, the dependence of the model on such terrain variables as curvature (curv), plan curvature (planc), profile curvature (profc), and solar radiation (solar) was stated; the terrain factors are shown in descending order by their importance for the model. The conclusion of the research is that soil salinity is largely affected by microrelief, convexity, or concavity, which in turn affects surface water retention.

In [30], the authors aim to monitor the aquifer of the eastern Nile Delta, Egypt, to check the level of salinity of groundwater occurring due to seawater intrusion (SWI); with this purpose, they propose to use the predictive regression model. The initial inputs were hydrogeological and hydrogeochemical data from the groundwater wells in 1996, 2007, and 2018, distance from the coastline, aquifer type, and hydraulic pressure. Using these input data, the authors constructed four baseline ML models allowing the forecasting of such indicators as the base exchange index (BEX), groundwater quality index for seawater intrusion (GQISWI), and water. The authors constructed such machine learning models as logistic regression, Gaussian process regression, feed-forward neural networks (FFNN), and one of the deep learning models, long short-term memory. The highest scores in terms of root mean square error (RMSE) and R^2 were received from the FFNN model, and testing showed $R^2 = 0.9667$ for BEX, $R^2 = 0.9316$ for GQISWI, and $R^2 = 0.9259$ for water quality. The ground samples of soils from arid areas in northwestern China and multispectral remote sensing data are used in [32]. The authors collected soil samples and determined the correlation between ground measurements and Landsat-8 OLI and Sentinel-2 MSI multispectral images. The results showed that the higher resolution Sentinel-2 MSI imagery gives more accurate results in estimating the salinity in desert areas.

The problem of predicting the location and degree of salinity of territories within the irrigation systems of the Waalhart and Bride rivers in South Africa was considered in [26]. The digital elevation models (DEMs) are considered as input data: DEM obtained from the satellite radar data (Shuttle Radar Topography Mission (SRTM)), and DEM based on photogrammetry (photogrammetrically-extracted digital surface model (DSM)). Areas with soil electrical conductivity (EC) values > 4 dS/m are indicated as soil with salinity. The salinity values based on soil samples were used to develop the model, to train the classifier, and to evaluate the score of the model. The binary classifier (based on the decision tree) demonstrated accuracy within 75%. The authors of the paper claim that the use of height data and their derivatives as input data for geo-statistics and machine learning has great potential for monitoring salt accumulation in the irrigated areas, especially for modeling subsurface conditions.

In [24], the authors considered the relationship between the backscattering coefficient (BC) and EC for Maha-Sarakham district in the northeast region of Thailand. ALOS-PALSAR provides data in four polarizations, HH, HV, VH and VV, with a resolution of 12.5 m. The total for the ground measurements is about 500 points from a depth of 5 cm, and the distance between each measurement is 20 m. Each measurement is identified by GPS with an accuracy of 4 m. The electrical conductivity was evaluated in the laboratory at 25 °C (dS m^{-1}). Several feed forward neural net models have been used to reconstruct the regression relationship between radar data and ground measurements. The models showed average values of the coefficient of determination $R^2 = 0.85$ and RMSE = 5.

A similar method for detecting soil salinity from Sentinel-1 radar data was described in [4]. The authors used a support vector regressor (SVR) to determine salinity, based on the field data.

An interesting approach was presented to solving the problem of salinity mapping using radar images and machine learning algorithms for the Mekong Delta were studied in [33]. The input values were the features generated from Sentinel-1 Synthetic Aperture Radar (SAR) C-band radar images, and the researchers considered the regression problem and predicted the electrical conductivity of soil. The generation of features from radar images was carried out using the method of constructing the Gray Level Co-occurrence Matrix (GLCM) proposed in [37]. The best result was achieved using the Gaussian Process Regressor (GPR) model with results of RMSE = 2.885, MAE = 1.897, and $R^2 = 0.808$.

Recently, the joint use of optical and radar data has gained great popularity. For example, ref [38] describes the combined use of Sentinel-1 and Sentinel-2 data, as well as a digital elevation model (DEM). In [7], the data from Sentinel-1 C radar and optical data from Sentinel-2 are used. Study [6] also uses Sentinel-1 SAR data and Sentinel-2 multispectral data to calculate the normalized differential salinity index (NDSI). The optical indices NDVI, SAVI, NDSI, SI based on Landsat 8 data and ALOS PALSAR-2 SAR data were used to form several combined optical radar indices (ORSDI) [3].

It should be noted that the list of input parameters for the developed soil salinity assessment models varies considerably. Insufficient or excessive number of parameters can lead to decreased quality of operation of machine learning algorithms. That is why it is necessary to optimize a set of model input parameters. In this case, the list of input parameters can be locally dependent. Correlation analysis, the importance of variables in projection (VIP), competitive adaptive sampling with re-weighting (CARS), the genetic algorithm (GA) [39–41], a method based on optimization of macro-parameters of the model and the number of input variables [42] are mentioned as means to solve this problem.

Summing up, it can be noted that algorithms based on decision trees, such as the above-mentioned decision tree [26], random forest (RF) [43,44], support vector machines (SVMs) [4,45], boosting models [38,46], and classical FFNN models [27,30,47], are quite often used as machine learning methods for predicting soil salinity.

According to the available information, the works devoted to the use of combined SAR and optical spectral bands data are limited to relatively small areas of the territory and to the application of a small number of machine learning models. However, a comparative analysis of the applicability of machine learning models and satellite data of different spectral ranges to substantially different territories, where both cultivated and uncultivated fields are located, is useful.

3. Method

In this research the authors have developed the approach proposed in [33], so that the SAR Sentinel-1 data were supplemented with terrain and temperature data, which, as noted in the literature sources [27,44] and as supported by this research (see Sections 4 and 5), improve the qualitative performance of the models. The authors also supplemented the SAR data with optical data, which improved the qualitative performance of the machine-learning models. In addition, there a comparative analysis of machine learning methods was performed and the capabilities were evaluated of the proposed salinity prediction technique for significantly different areas of southern Kazakhstan's territory.

The flowchart of the process of salinity detection is shown in Figure 1. The process of training and applying the model consists of the following steps:

1. Obtaining the salinity data using the field studies;
2. Obtaining the radar, multispectral and SRTM data from Google Earth Engine;
3. Extraction of linear back scattering intensity in VV and VH polarizations;
4. Texture analysis using the GLCM method;
5. Application of the machine learning algorithms and evaluation of the quality of the trained model;
6. Mapping the selected areas of the territory.

Multispectral data from Landsat-8 were obtained and used to calculate the spectral indices.

As a result, as will be shown below, the regressors that demonstrate the best results were identified, and limitations in the application of the method for individual subsets of data were determined.
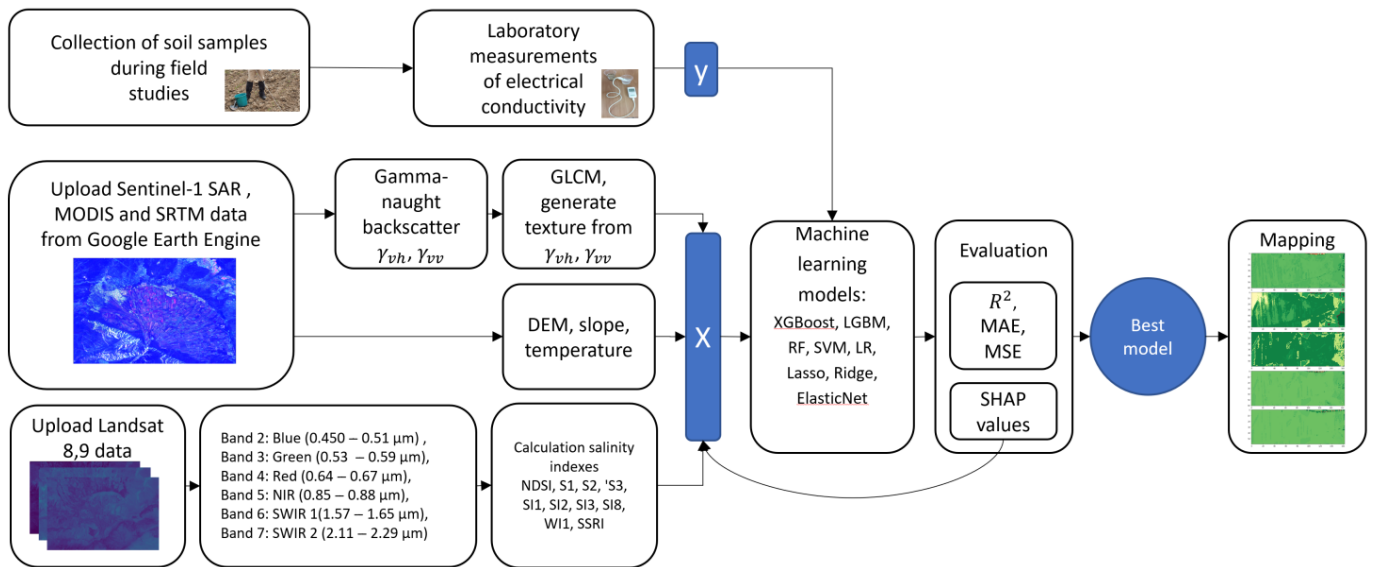
Collection of soil samples during field studies

Laboratory measurements of electrical conductivity

y

Upload Sentinel-1 SAR , MODIS and SRTM data from Google Earth Engine

Gamma-naught backscatter γ_{vh}, γ_{vv}

GLCM, generate texture from γ_{vh}, γ_{vv}

DEM, slope, temperature

X

Machine learning models: XGBoost, LGBM, RF, SVM, LR, Lasso, Ridge, ElasticNet

Evaluation
R^2, MAE, MSE
SHAP values

Best model

Mapping

Upload Landsat 8,9 data

Band 2: Blue (0.450 − 0.51 μm) ,
Band 3: Green (0.53 − 0.59 μm),
Band 4: Red (0.64 − 0.67 μm),
Band 5: NIR (0.85 − 0.88 μm),
Band 6: SWIR 1(1.57 − 1.65 μm),
Band 7: SWIR 2 (2.11 − 2.29 μm)

Calculation salinity indexes NDSI, S1, S2, 'S3, SI1, SI2, SI3, SI8, WI1, SSRI

Figure 1. Methodological scheme of the study. X—input variables(features), y—target value.

3.1. Data Preparation

Field studies of the soil were carried out during the expeditions on 23 May (Shelek), 26 May, 17, 18 and 29 June (Kapchagay) and 18 July 2022 (Alakol). The sampling route was designed taking into account the availability of vehicles and the expected presence of pronounced areas of saline soil. A portable GPS device (Garmin 65) with a positioning accuracy of ≤5 m was used to record the geographic coordinates. Although the positioning accuracy level of the GPS device was not perfect, it was sufficient to ensure reasonable position matching between the sampling site and satellite image pixels (since 10 m is the size of 1 image pixel; in other words, a 10 m resolution for the Sentinel 1 satellite). The sampling locations were photographed and described.

A total of 207 soil samples were collected in the area of Lake Alakol, near Kapchagay Reservoir and Shelek village. Accordingly, three subsets of soil samples were formed: Alakol (45 samples), Kapchagay (84 samples), and Shelek (78 samples). Figure 2 shows the locations of soil sample collection near the Kapchagay reservoir and on the foothill plains near Lake Alakol.
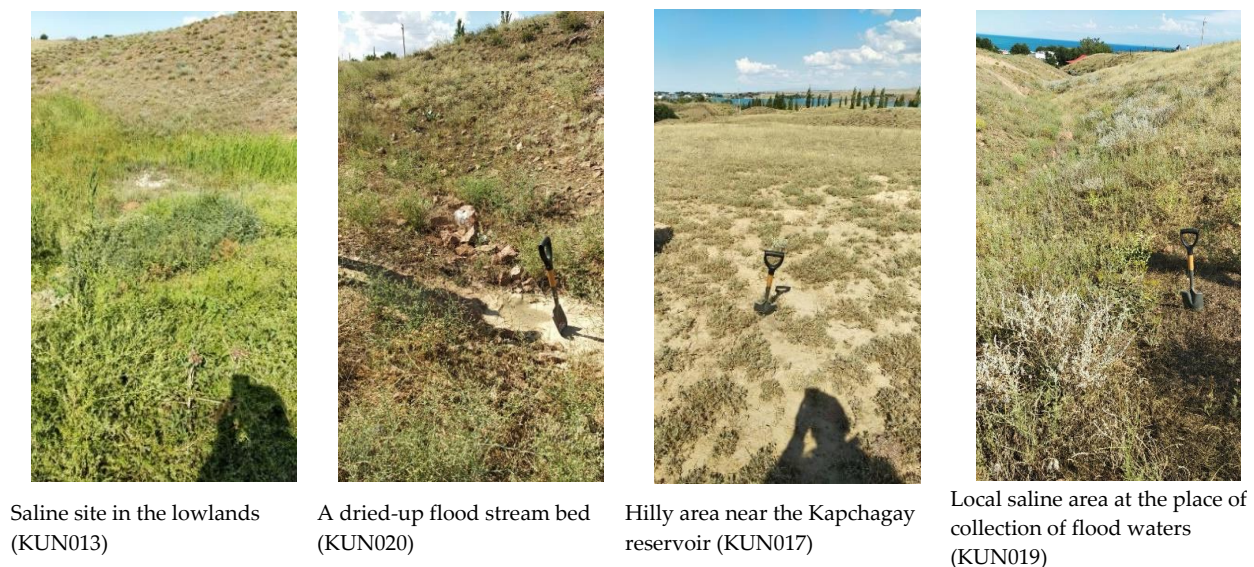
Saline site in the lowlands (KUN013)

A dried-up flood stream bed (KUN020)

Hilly area near the Kapchagay reservoir (KUN017)

Local saline area at the place of collection of flood waters (KUN019)

Figure 2. *Cont.*

| Foothill plain in the area of Lake Alakol (ALK8) | Saline area on the plains near Lake Alakol (ALK30) | Foothill plain in the area of Lake Alakol (ALK25) | A field sown with alfalfa near Lake Alakol (ALK17) |

Figure 2. Soil sampling locations near Kapchagai Reservoir and Lake Alakol.

Figure 3 shows soil sampling routes. The collection conditions showed significant differences. Shelek samples were collected on moist loamy arable land with corn plantations at the end of May. At the same time, samples of Kapchagay were collected on dried hilly sandy soil mainly outside of ploughed areas. Alacol samples were collected in mixed mode both within and outside of the arable areas. As will be shown below, this had a significant impact on the quality of the regression analysis. All soil samples were properly sealed, labeled, and transported to the laboratory for measurement of their electrical conductivity.
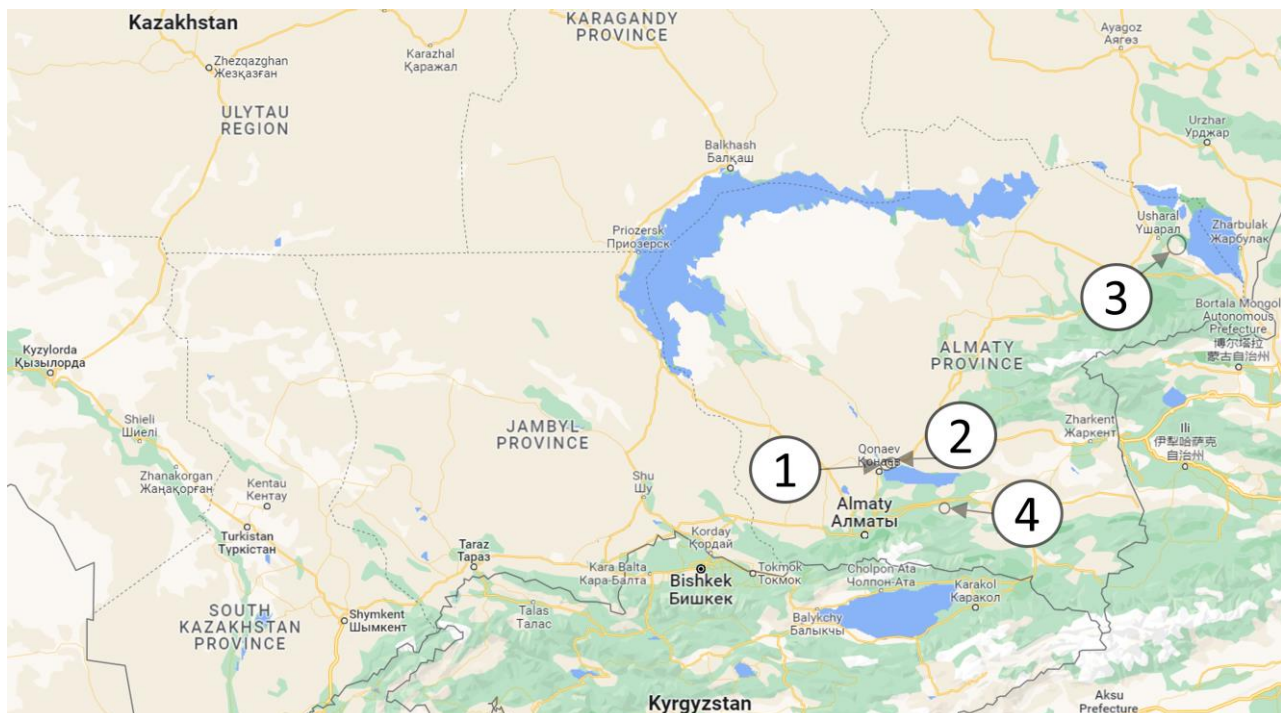

Figure 3. Map of soil sample collection in Almaty and Zhetysu regions (1, 2—Kapchagay, 3—Alacol, 4—Shelek).

The samples were completely dried and sieved through a 2-mm mesh sieve to remove vegetation and stone residues. The prepared soil samples were mixed with water at a ratio of 1:5 (one weight fraction of soil and five weight fractions of water). The obtained solutions were left to settle for one day for complete dissolution of the fractions. Then, the

electrical conductivity of the soil was determined using a digital meter (Hanna GroLine HI9814) at room temperature (25 °C).

The data from Sentinel-1 Synthetic Aperture Radar (SAR) and optical data obtained in April–August 2022 were used as initial data. Figure 4 shows the images synthesized from the SAR data, which are composed of a combination of colors imitating the reflected radar signals in the following way: Red is VH, Green is VV, and Blue is VH/VV. The first letter of the signal code indicates the polarization of the emitted signal and the second letter the polarization of the reflected signal.
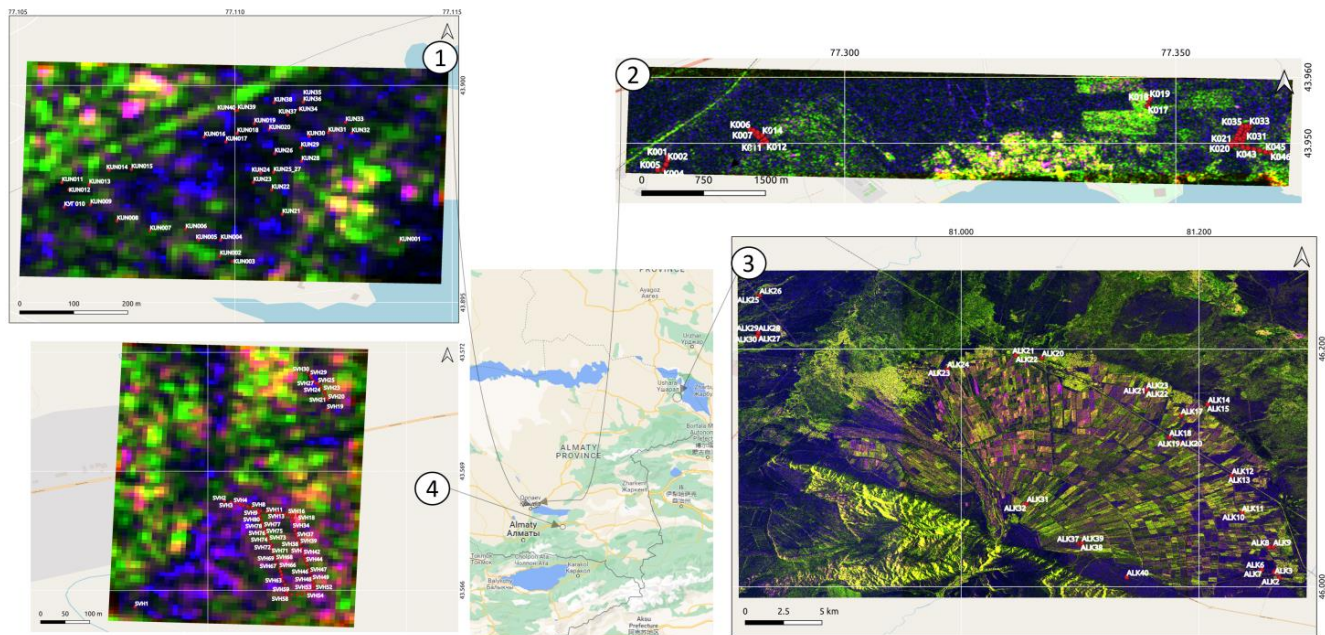

Figure 4. The color composite image of the SAR Sentinel-1 data from Alakol lake, Kapchagay reservoir and Shelek region (Red: VH, Green: VV, Blue: VH/VV). 1, 2—Kapchagay, 3—Alakol, 4—Shelek.

The pre-processing of Sentinel 1 SAR radar data resulted in a data set, a fragment of which is shown in Figure 5. The figure demonstrates the set obtained from the SAR data in April 2022. The entire dataset can be downloaded from the link: https://www.dropbox.com/sh/djlm3ox8briepv3/AAD691Fa2hPGtLIl5IEZ1Urka?dl=0 (accessed on 1 August 2023).

Domain	DomNumber	elco50	Long_dec	Lat_dec	dem	slope	temp	dissimilarity_vv_l	contrast_vv_l	homogeneity_vv_l	ASM_vv_l	energy_vv_l	correlation_vv_l	entropy_vv_l	dissimilarity_vh_l	contrast_vh_l	homogeneity_vh_l	ASM_vh_l	energy_vh_l	correlation_vh_l	entropy_vh_l	gamma_vv_l	gamma_vh_l
Shelek	1	0.34	78.09316	43.56558	569	0.5417	25.29	40.8	2366.2	0.0052130	0.05	0.2236	0.879244	4.643856	52.7	3390	0.00238237	0.05	0.22360679	0.159040	4.5638	0.2321	0.01293
Shelek	1	0.37	78.09541	43.56825	567	0.4361	25.29	30.549	1456.2499	0.0780879	0.049	0.2236	0.821834	4.563856	51.1499	3892.049	0.00409354	0.049	0.22360679	0.185448	4.5638	0.0869	0.01199
Shelek	1	0.74	78.09569	43.56819	567	0.4361	25.29	37.85	2367.85	0.0531232	0.05	0.2236	0.709814	4.563856	53.05	4246.65	0.00419974	0.05	0.22360679	0.451169	4.5638	0.0624	0.02288
Shelek	1	0.82	78.09580	43.56816	567	0.4361	25.29	42.45	2665.55	0.0524731	0.05	0.2236	0.524217	4.403856	59	4648	0.00263562	0.05	0.22360679	0.372973	4.5638	0.1562	0.02942
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Kapch	2	0.05	77.27322	43.94797	567	1.2011	40.41	33.65	1873.65	0.0034603	0.05	0.2236	0.516920	4.403856	43.5499	2896.15	0.00959691	0.049	0.22360679	0.412331	4.5638	0.0054	0.00367
Kapch	2	0.06	77.27326	43.94744	567	1.2011	40.41	59.199	4911.5999	0.0046823	0.049	0.2236	0.335806	4.643856	39.05	2529.55	0.03893189	0.05	0.22360679	0.796284	4.5638	0.0281	0.00247
Kapch	2	0.05	77.27299	43.94680	555	1.646	40.76	31.25	1574.55	0.0377267	0.05	0.2236	0.600938	4.483856	44.85	2989.85	0.03106235	0.05	0.22360679	0.433423	4.5638	0.0076	0.00220
Kapch	2	0.04	77.27265	43.94598	555	1.646	40.76	33.75	2008.15	0.0142133	0.05	0.2236	0.707066	4.563856	41.65	2490.95	0.02800886	0.05	0.22360679	0.508782	4.6438	0.0066	0.00093
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Alacol	3	0.53	81.09905	46.03648	492	0.5497	41.56	51.1	3834.4	0.0109399	0.05	0.2236	-0.134384	4.483856	47.8	3364.6	0.00324971	0.05	0.22360679	0.375423	4.5638	0.0301	0.01251
Alacol	3	0.47	81.10058	46.03644	492	0.5497	40.5	25.399	1035.5999	0.0331120	0.049	0.2236	0.721583	4.403856	38.35	2415.65	0.01014350	0.049	0.22360679	0.223943	4.5638	0.1380	0.01110
Alacol	3	0.38	81.10148	46.03686	492	0.5497	40.5	35.75	1906.25	0.0287192	0.05	0.2236	0.568201	4.483856	42.6	2336.5	0.00350400	0.049	0.22360679	0.682572	4.5638	0.0547	0.00155
Alacol	3	0.34	81.13923	46.00802	487	3.223	45.03	51.75	3656.25	0.0034962	0.05	0.2236	0.570489	4.563856	40.8	2775.3	0.00426183	0.05	0.22360679	0.648735	4.6438	0.0134	0.00508

Figure 5. Part of the Santinel-1 SAR dataset on the salinity of certain areas of South Kazakhstan.

The radar dataset consists of field soil samples (electrical conductivity—elco50), 16 features that were generated using a radar image and the GLCM method, temperature and information about the terrain (dem, slope). Table 1 contains abbreviations and

descriptions of the dataset. For radar data augmentation, the measured altitude by GPS, the slope computed with the use of digital elevation model, and temperature data, which can affect the salinity classification model, were proposed. USGS MODIS Earth Surface Temperature or land surface temperature (LST) and ground elevation model (ELV) data were used to extend the data set. The resolution of the images received during the process were as follows: 1 km for LST and 30 m for ELV. Height above sea level and slope of relief, or angle of land, were received from ELV data.

Table 1. Names of features and their description.

Name	Description
	Target value
elco50	Soil salinity, field data
	Features generated using SAR Sentinel-1 data
dissimilarity_vv	Dissimilarity of gray level co-occurrence matrix for polarization VV
contrast_vv	Contrast of gray level co-occurrence matrix for polarization VV
homogeneity_vv	Homogeneity of gray level co-occurrence matrix for polarization VV
energy_vv	Energy of gray level co-occurrence matrix for polarization VV
entropy_vv	Entropy of gray level co-occurrence matrix for polarization VV
gamma_vh	Linear backscatter intensity in VV polarization
gamma_vv	Linear backscatter intensity in VH polarization
dissimilarity_vh	Dissimilarity of gray level co-occurrence matrix for polarization VH
contrast_vh	Contrast of gray level co-occurrence matrix for polarization VH
homogeneity_vh	Homogeneity of gray level co-occurrence matrix for polarization VH
energy_vh	Energy of gray level co-occurrence matrix for polarization VH
entropy_vh	Entropy of gray level co-occurrence matrix for polarization VH
correlation_vv	Correlation VV
correlation_vh	Correlation VH
ASM_vh	Angular second moment VH
ASM_vv	Angular second moment VV
	Environmental features
Long_dec	Longitude in decimal coordinates WGS84
Lat_dec	Latitude in decimal coordinates WGS84
Altitude	Measured of altitude by GPS
temp	MODIS land surface temperature
slope	Calculated slope from DEM
	Spectral indexes (see Table 2)

Table 2. Spectral indices used as input for machine learning models.

Spectral Indexes	Ref.
$NDSI = \frac{red - nir}{red + nir}$	[48]
$S1 = \frac{blue}{red}$	[49]
$S2 = \frac{blue - red}{blue + red}$	[49]
$S3 = \frac{green \times red}{blue}$	[49]

Table 2. *Cont.*

Spectral Indexes	Ref.
$SI1 = \sqrt[2]{green \times red}$	[50]
$SI2 = \sqrt[2]{green^2 + red^2 + nir^2}$	[51]
$SI3 = \sqrt[2]{green^2 + nir^2}$	[52]
$SI8 = \frac{blue \times red}{green}$	[53]
$WI1 = 0.1761 \times \text{green} + 0.322 \times \text{red} + 0.3396 \times \text{nir}$	[54]
$SSRI = \frac{nir}{\sqrt[2]{green \times red}}$	[55]
$NDSIre = \frac{red - swir\_16}{red + swir\_16}$	*
$SI3re = \sqrt[2]{green^2 + swir\_16^2}$	*
$SSRIre = \frac{swir\_16}{\sqrt[2]{green \times red}}$	*

The following spectral ranges were used to calculate the indices [56]: Band 2, Blue (0.450–0.51 μm), Band 3, Green (0.53–0.59 μm), Band 4, Red (0.64–0.67 μm), Band 5, Near-Infrared (0.85–0.88 μm), Band 6, SWIR 16 (1.57–1.65 μm), Band 7, SWIR 22 (2.11–2.29 μm). To use the "swir_16" range data, additional spectral indices are proposed using this range instead of "nir" (marked with "*").

The data from Landsat-8 satellite with 30 m. resolution were used to calculate the spectral indices listed in Table 2.

3.2. Analysis of Data

To train the machine learning models, an approximately equal distribution of objects with different salinity values is desirable. For this purpose, the target column was converted into a column of five salinity classes for loam soil according to [57,58]. Table 3 shows the distribution of samples by salinity class.

Table 3. Distribution of field data by salinity class based on electrical conductivity as $EC_{1:5}$ for loam soil.

Salinity Class	Class Number	Number of Samples	$EC_{1:5}$ Range for Loams (dS/m)
Non-saline	0	60	0–0.18
Slightly saline	1	42	0.19–0.36
Moderately saline	2	41	0.37–0.72
Highly saline	3	21	0.73–1.45
Severely saline	4	43	>1.45

Due to the fact that the distribution of samples by salinity class was approximately equal (except for class 3—high saline), it was decided to divide the data set into training and test randomly, using the standard utility train_test_split Sklearn [59].

An analysis of the correlation of the input parameters of the SAR data shows a significant relationship with the subset of the analyzed data (Figure 6). For example, in the Alakol dataset, ASM_Vh has a zero value, while the rest of the data sets demonstrate an almost 100% correlation with energy_vh. There is a positive correlation between dem, temp and slope in the Alakol and Kapshagai sets, while in the Shelek set there is a significant negative one. The correlation between elco50 (target value) and coordinates also differs significantly for different datasets, including the sign of the correlation.
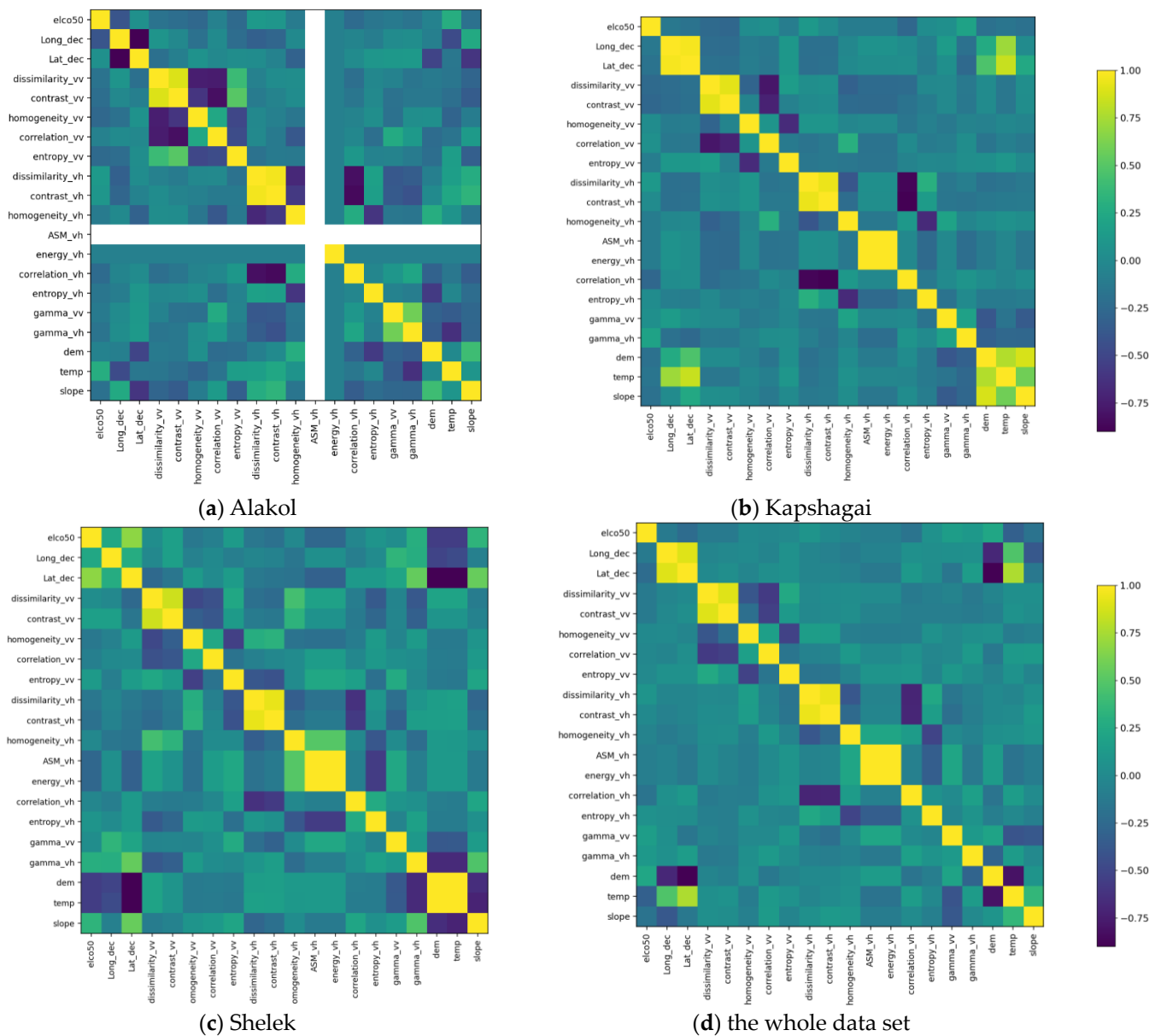
(**a**) Alakol

(**b**) Kapshagai

(**c**) Shelek

(**d**) the whole data set

Figure 6. Feature correlation heatmaps.

In this regard, it was decided to evaluate the performance of the machine learning models on the Alakol, Kapchagay, and Shelek data subsets separately.

It should be noted that some input parameters can be redundant and even reduce the quality of the model performance. To optimize the set of properties, random search algorithms are used, for example, the genetic algorithm [7,60]. The disadvantage of this approach consists of rather significant computational costs.

The second question is to evaluate the influence of input parameters (features) on the model results. In other words, this is required to explain the result obtained by the model. Model specific and model agnostic methods of machine learning interpreting methods are used to solve this problem [61]. Due to the fact that the authors analyze the machine learning models based on different mathematical principles, it is convenient to apply an agnostic interpretation model, such as SHapley Additive exPlanations (SHAP) [62]. The authors also used the same model, as will be described below, to optimize multiple features.

3.3. Machine Learning Models

The authors previously considered the possibility of applying a wide range of machine learning algorithms, both classical and modern [63]. Obviously, the application of deep learning models could give a good result. However, the generated dataset is relatively small. The authors are also not aware of similar initial SAR datasets of significant volume or pre-trained models of deep neural networks to solve the salinity estimation problem, which would allow the application of transfer learning techniques for its tuning. Therefore, the use of deep learning models was considered inappropriate. The authors used Busting models [64], Support vector machines, and classical regression algorithms as a reference point (base line) for comparative analysis (Table 4).

Table 4. Machine Learning Models.

Regression Model	Abbreviation	Method	References
XGBoost	XGB	Ensemble learning method based the gradient boosted trees algorithm.	[65]
LightGBM	LGBM	Ensemble learning method based the gradient boosted trees algorithm.	[66–68]
Random forest	RF	Ensemble learning method based on bagging technique	[69]
Support vector machines	SVM	Linear and non-linear classification based on the technique named kernel trick	[70]
Linear regression	LR	Linear approach to modeling impact of independent variables to dependent value or target variable.	[71]
Lasso regression	Lasso	Based on the use of such a regularization mechanism that not only helps in reducing overfitting but it can help in feature selection.	[72]
Ridge regression	Ridge	A regularization mechanism is used to prevent over-training (overfitting).	[73,74]
Elastic net	ElasticNet	Hybrid of ridge regression and lasso regularization	[75]

To assess the quality of regression models, the following accuracy indicators are often used [31]: coefficient of determination (R^2), Mean squared error (MSE) and Mean absolute error (MAE). The quality indicators used to estimate the regressors, as well as the corresponding expressions, are listed in Table 5.

Table 5. Quality metrics for regression models.

Accuracy Index	Abbreviation	Equation	Explanation
Determination coefficient	$R^2/r2\_score$	$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ $SS_{res} = \sum_{i=1}^{m_k} \left(y^{(i)} - h^{(i)}\right)^2$ $SS_{tot} = \sum_{i=1}^{m_k} \left(y^{(i)} - \overline{y}\right)^2, \overline{y} = \frac{1}{m_k} \sum_{i=1}^{m_k} y^{(i)},$	where $y^{(i)}$ is the actual value; $h^{(i)}$ is the estimated value (the value of the hypothesis function) for the i-*th* sample; $m_k \in m$ is a part of the training sample (the set of labeled objects)
Mean Absolute Error	MAE	$MAE = \frac{\sum_{i=1}^{n} \left|y^{(i)} - h^{(i)}\right|}{n}$	where *n* is a simple size; when evaluating the performance of the model on the test set *n* is the size of the test set
Mean squared error	MSE	$MSE = \frac{\sum_{i=1}^{n} \left(y^{(i)} - h^{(i)}\right)^2}{n}$	

4. Results

4.1. Evaluation of Regression Models

The quality of performance of machine learning models was evaluated by Random permutations cross-validation (ShuffleSplit). In this case, the raw data are divided into training and test sets randomly in a given proportion (in this case, 80% is training data, and 20% is test data). To ensure the statistical significance of the result, such splitting was performed 200 times for each regressor model. The obtained values of the estimates were averaged and variance was calculated for them using the statistics library. The results of the models are shown in Tables 6 and 7 (full results of the calculations are presented in Appendices A and B), where varMAE, varMSE, VarR^2 are the variance of the obtained estimates, and Duration is the training and estimation time of the model in seconds.

Table 6. Results of machine learning models using SAR data.

Dataset	Regression Model	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Full Dataset	XGB	0.644	1.991	**0.282**	0.024	3.542	0.046	28.14647
	RF	0.738	2.318	0.093	0.026	3.956	0.131	28.71924
	LR	0.913	2.515	−0.02	0.027	3.805	0.033	0.406411
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.380981
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.355016
	LGBM	0.814	2.178	0.183	0.033	3.721	0.029	11.27983
	Ridge	0.884	2.421	0.044	0.023	3.849	0.023	1.218743
	SVM	0.791	2.27	0.139	0.031	3.681	0.018	0.627324

Table 7. Results of machine learning models using radar and optical data.

Dataset	Regression Model	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Full Dataset	XGB	0.569	1.889	**0.339**	0.023	3.49	0.057	28.16568
	RF	0.692	2.349	0.058	0.032	3.925	0.276	39.8665
	LR	0.923	2.43	−0.095	0.023	2.712	0.184	0.409904
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.375001
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.338127
	LGBM	0.69	1.935	0.325	0.03	3.707	0.048	11.27485
	Ridge	0.795	2.211	0.145	0.02	3.66	0.055	4.66742
	SVM	0.811	2.242	0.145	0.029	3.573	0.015	0.791433

The XGB regressor shows the best results (highlighted in bold). Figure 7 shows a scatterplot of the salinity measured and predicted by XGB for different data sets. The diagonal line in the figure shows the optimum line, where the prediction value coincides with the actual value.
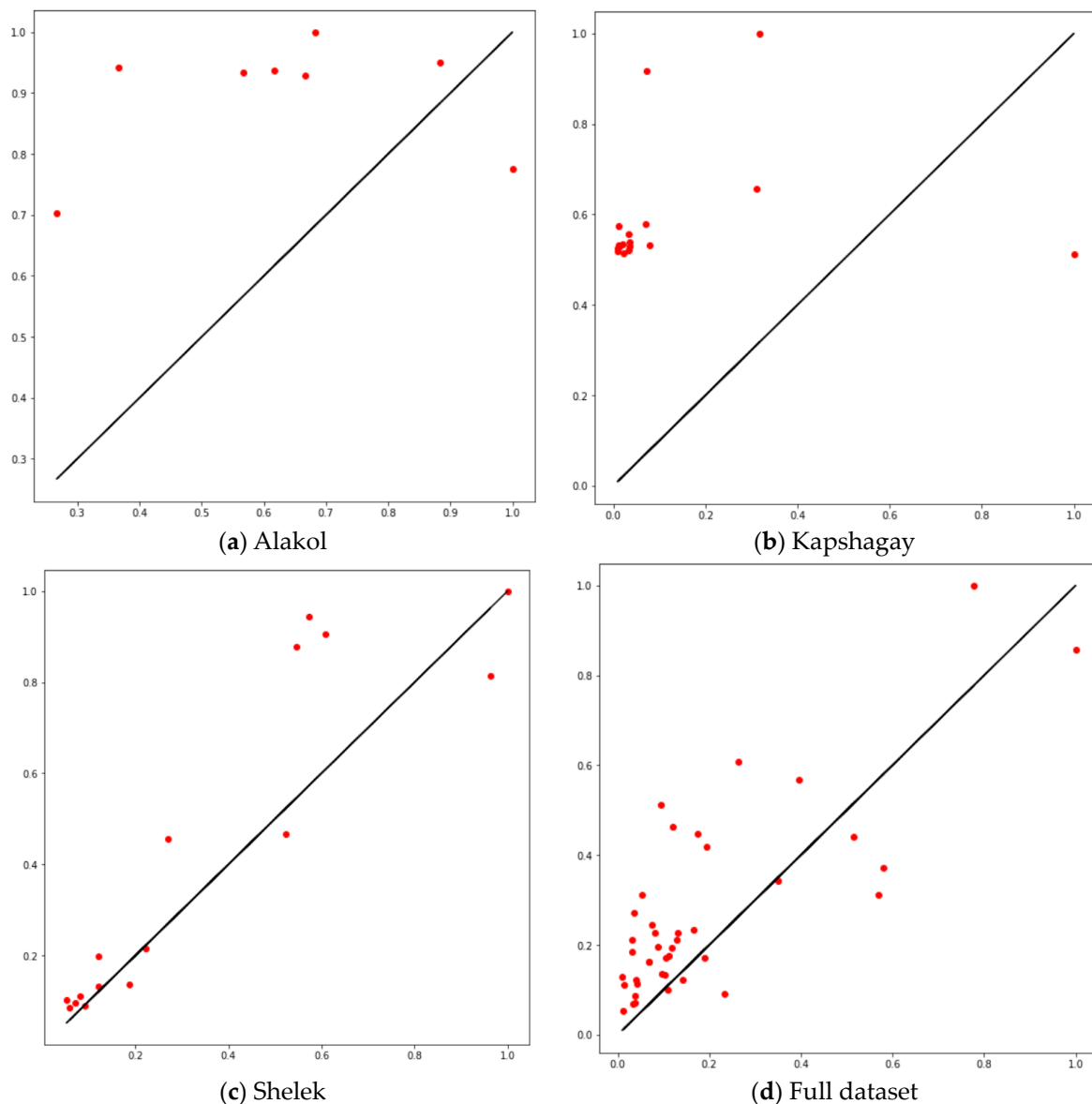
(**a**) Alakol

(**b**) Kapshagay

(**c**) Shelek

(**d**) Full dataset

Figure 7. Scatterplots of laboratory and predicted salinity values for different data sets.

It can be seen that the results of the regressor on the Shelek dataset have the point density closest to the optimum line. The experiments with this dataset show that $R^2 = 0.473$ using SAR data and $R^2 = 0.654$ for the full set of input variables including radar, spectral indices, temperature, and terrain information. Without going into detail, it should be noted that the classification of data using the XGB Classificator with SAR data for five salinity classes gave an average accuracy of 54%. Applying the same classifier to calculate three and two classes gives an average accuracy of 65% and 75%, respectively. At the same time, the use of both spectral indices and radar data gives slightly better average estimates for classes 5, 3 and 2: 58%, 71%, 77%, respectively.

4.2. Analysis of Influence of Input Parameters

To analyze the dependence of the model results on the input parameters, the authors used the SHAP library [62]. The results of the analysis of the XGB model are shown in Figure 8. In the figure, a, c—the influence of the parameters obtained with SAR data; b, d—the influence of spectral data and optical indices.
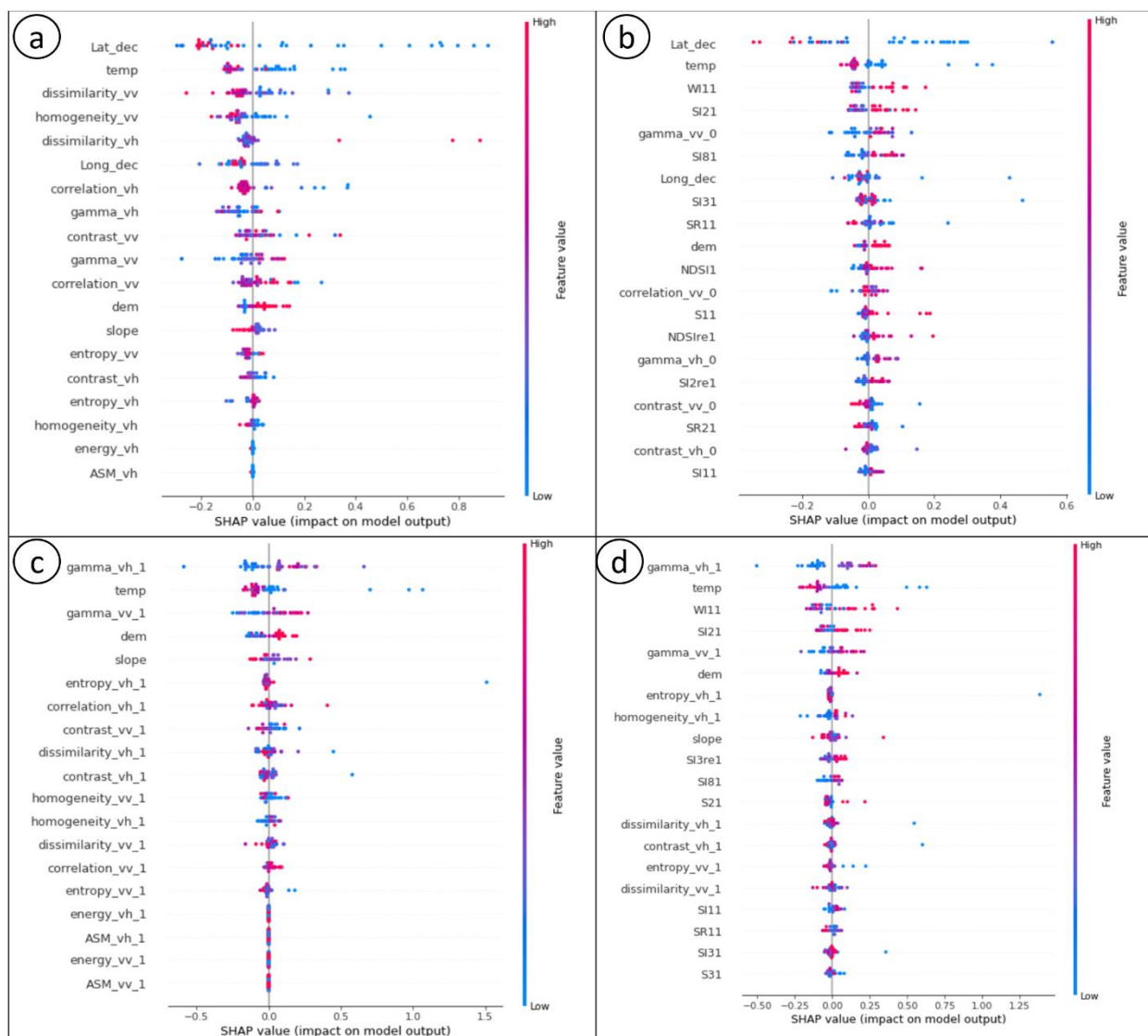
Figure 8. Impact of input values on model output.

The regression result depends on almost all input parameters. For example, a decrease of latitude positively affects the salinity value; in other words, the more southern the considered zone, the more likely is high salinity value; high slope value, on the contrary, negatively influences the soil salinity degree—high salinity on a slope is improbable. Temperature is the second most influential parameter. When using the optical data, the third and fourth most important factors are optical indices.

Due to the significant geographical dispersion of the data sets, the use of coordinate values in the number of input parameters may not be appropriate. To this end, in Figure 8, fragments c, d show the effect of parameters without regard to coordinates.

The obtained results allow us to range the input parameters for the regression model. The maximal influence has the parameter gamma_vh. The second parameter in the rank is the temperature: the higher the temperature results, the higher the value of the electrical conductivity (salinity) of the soil. The third and fourth parameters are the optical indices, water index WI, and salinity index SI2 (Figure 8d): the higher the average WI and SI2 values, the higher the salinity.

It can be assumed that the exclusion of terrain and temperature data from the input parameters will reduce the quality of the regression models, which is confirmed by the results of the computational experiment (see Table 8).

Table 8. Results of regressors performance with the SAR dataset, in which the parameters slope, dem, temp are excluded.

Dataset	Regressor	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Full Dataset	XGB	0.677	2.053	**0.227**	0.027	3.329	0.034	29.84913
	RF	0.744	2.445	-0.117	0.026	3.258	0.444	29.09285
	LR	0.919	2.309	-0.017	0.016	2.727	0.072	0.403919
	Lasso	0.969	2.449	-0.047	0.021	3.032	0.011	0.383972
	ElasticNet	0.969	2.449	-0.047	0.021	3.032	0.011	0.347101
	LGBM	0.763	2.029	0.205	0.026	3.131	0.032	10.7253
	Ridge	0.881	2.274	0.049	0.021	2.897	0.019	1.139951
	SVM	0.747	2.117	0.183	0.03	3.279	0.034	0.633299

At the same time, the set of input parameters can be reduced without degrading the quality of the regressors. However, it must be taken into account that the relationship between the input parameters and the target variable is not linear and the input parameters affect the result together. Therefore, the simple removal of "insignificant" parameters can lead to deterioration of regression indicators. In order to remove the parameters that reduce the performance of the regression model, the authors used the following method. It is known that SHAP values form an objective relation between the model results and the input parameters. Consequently, it is possible to exclude from the properties those variables that do not affect the result and those that have a strong correlation between them. Figure 9 shows the correlation matrices of the original (left) and reduced (right) sets of SHAP values.
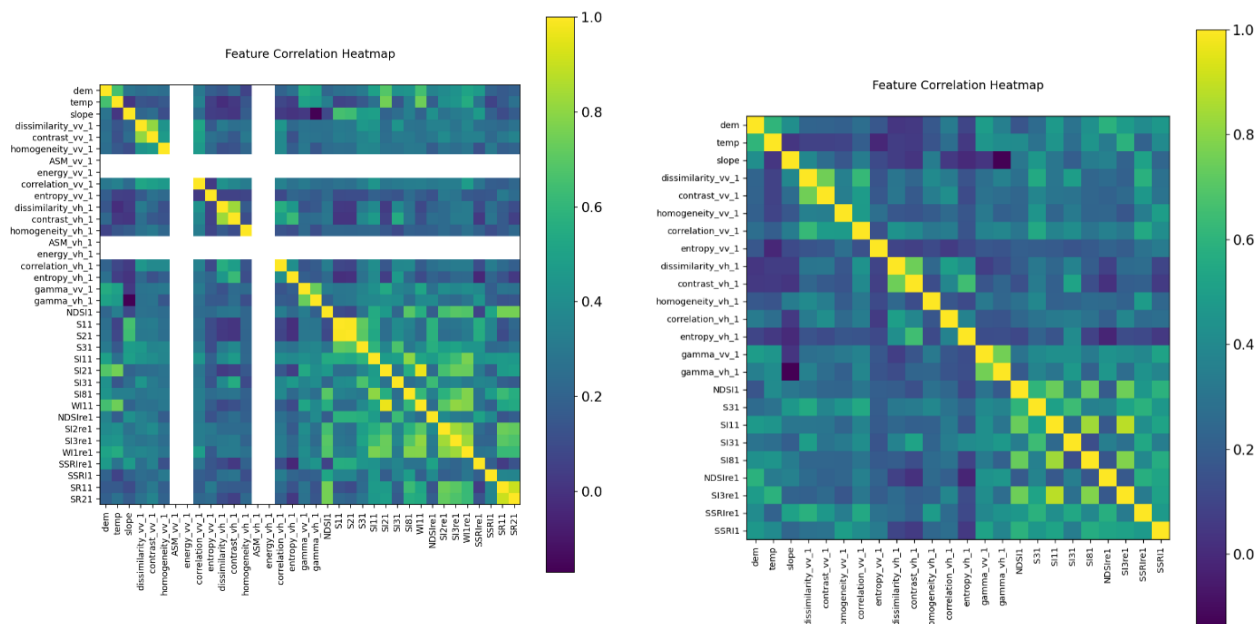

Figure 9. Correlation matrix of SHAP values.

After the above-described optimization, the models use 24 input parameters instead of 36 and their results have improved, especially for XGB, RF, Ridge, and SVR (see Table 9).

Table 9. Results of machine learning models performance after optimizing the number of input parameters.

Dataset	Regressor	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Full Dataset. The set of features are optimized.	XGB	0.575	1.858	**0.356**	0.023	3.526	0.061	32.16551
	RF	0.681	2.242	0.138	0.027	3.877	0.128	32.89764
	LR	0.88	2.412	−0.065	0.021	3.024	0.225	0.386937
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.351073
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.335103
	LGBM	0.694	1.929	0.33	0.032	3.766	0.048	10.12191
	Ridge	0.764	2.104	0.221	0.02	3.671	0.033	1.68862
	SVM	0.71	2.053	0.273	0.029	3.786	0.034	0.608407

The results of the regressors' performance with an exclusively optimized list of spectral indices as input are shown in Appendices C and D. The list of input parameters was reduced from 20 to 16 and the results improved by about 1%.

5. Discussion

To evaluate the obtained results, in addition to the numerical metrics, it is useful to compare them with the results previously obtained for a given territory with employment of another model. Figure 10 shows the results of applying the XGB regressor to the salinity mapping at a site approximately corresponding to site 4 (Shelek) in Figure 3. The model was trained using the spectral indices. The figure shows a section of the terrain in the foothills of the Zailiysky Alatau (1) and a satellite image Landsat 8 (band 4, red) (2). The sites of soil sample collection are marked on the map (1) by green rectangles. The right part (3) shows the model of predicted salinity levels obtained using Landsat 8, 9 images from 1, 9, 10, 17 and 26 April 2022 (from top to bottom). Green areas correspond to the minimum level of salinity, and lighter areas, going to yellow and red, to a high level of salinity. Note that the cloudy days on 9 and 10 April distorted the result. The bottom right (4) shows the results of the prediction using the RF-based model from [44] based on temperature data (hereafter Mtemp). In this case, it can be seen that the Mtemp model predicts two classes of soil salinity, which in general do not coincide with the measured electrical conductivity values of soil samples and XGB regressor predictions.

Figure 11 illustrates the dynamics of forecasting changes in the condition of the soil cover in the areas of Zailiisky Alatau on 1 April (2) and 23 August (3), 2022. The zone for collecting the soil samples (on 23 May 2022) is marked on the map (1) by a white rectangle. This collection area corresponds to Figure 10. The map in Figure 11 shows approximately 25 times the area of the Zailiiskiy Alatau.

Figure 12 shows the simulation results for Alakol Lake area in comparison with Mtemp model. It can be seen that the models again show significantly different results.

In general, the computational experiments have shown that the satellite SAR data can be used to estimate the salinity in the southern regions of Kazakhstan. Expanding the number of input variables by a set of spectral indices leads to improved results. It can be stated that:

1. XGBRegressor has the best quality indicators for the considered regressors; LightGBM is the second in terms of quality indicators.
2. The results of the constructed regression are significantly better on moist cultivated soil (Shelek) than on the entire data set.
3. The quality of work on the datasets from Alakol and Kapchagay is low. It can be assumed that sampling in a local area of hilly terrain (Kapchagay) and large sampling areas in the Alakol region require a more laborious process of soil data generation, for example, in the form of five-spot sampling [38].

4. Comparison of the results of the XGB regressor with the Mtemp model shows that the models produce significantly different results. It can be assumed that a possible reason for the discrepancy is that Mtemp was trained without using data from the regions of Kazakhstan.
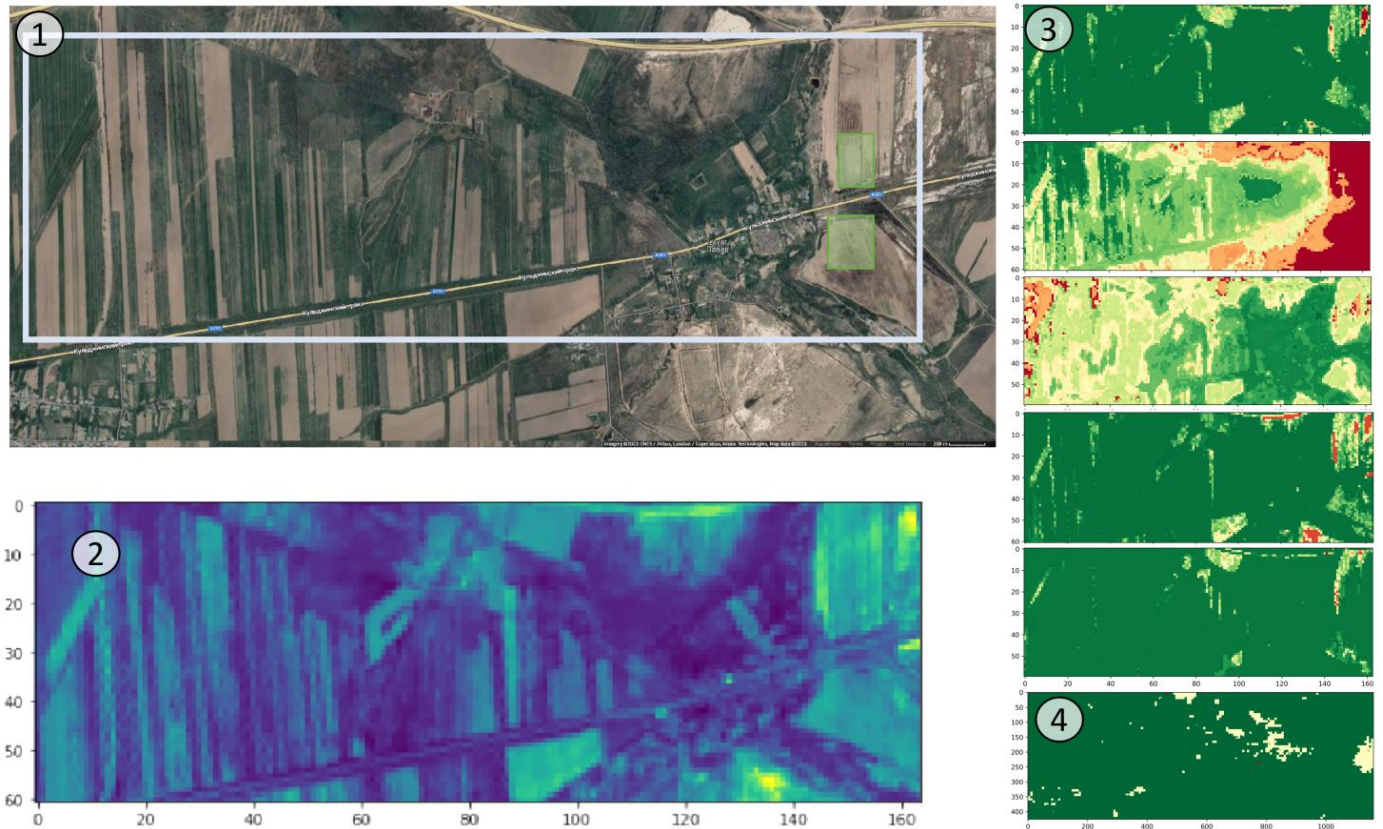

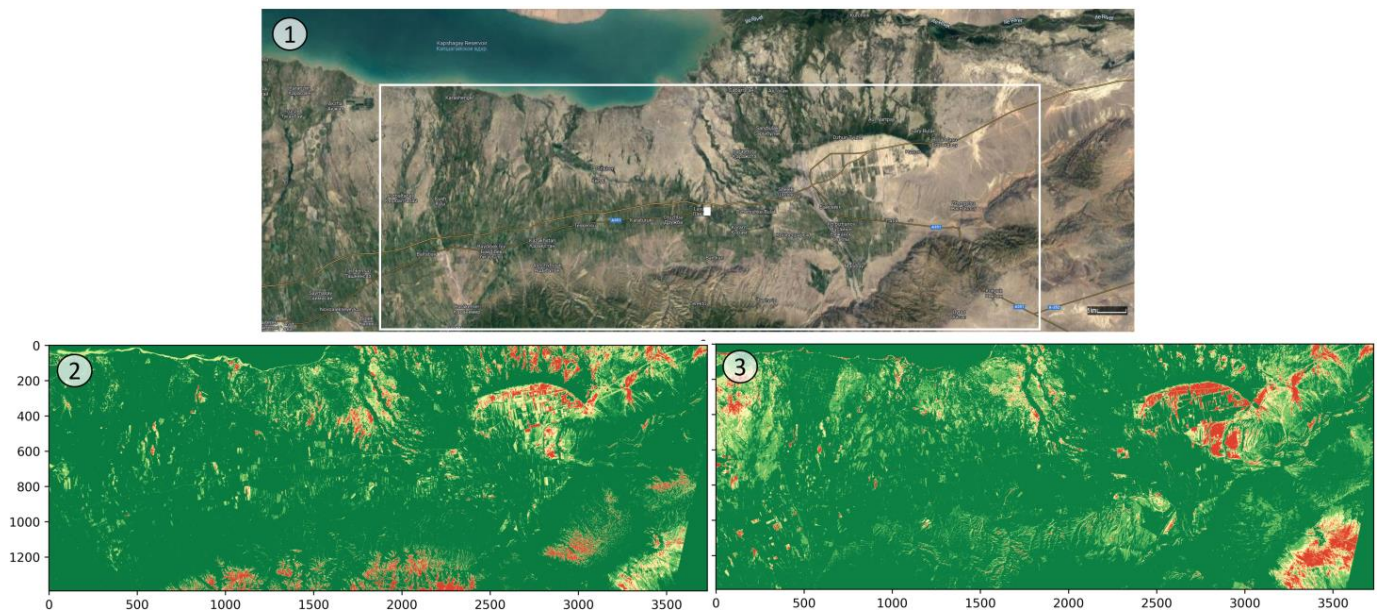**Figure 10.** Mapping of soil salinity in April 2022 in Shelek area.

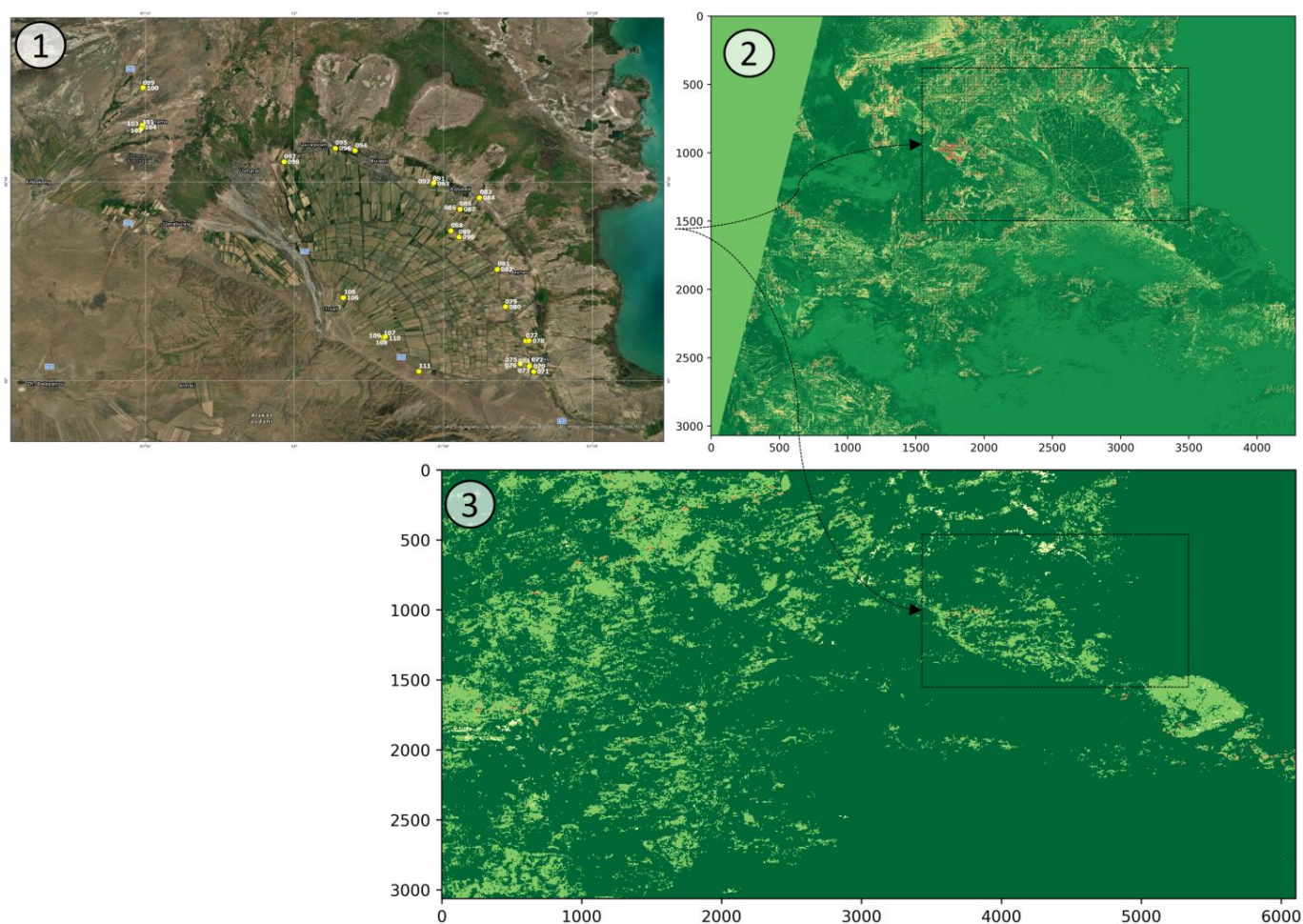**Figure 11.** Dynamics of soil condition changes between the southern bank of the Kapchagai reservoir and the foothills of the Zailiisky Alatau.

Figure 12. Soil salinity in the area of Lake Alakol. Top left (**1**) is the field study area. Top right (**2**) is predicted by the developed model. Bottom (**3**) is predicted by the Mtemp model.

The resulting regression parameters are relatively worse compared to those given in articles by other authors [76]. It can be assumed that this is a consequence of the complexity of the foothill landscape and the large difference in time when performing ground studies. For example, soil samples from the Kapchagay dataset were collected on dry hilly terrain, where there were only small (3–5 m) local areas of salinity that were not visible from space. Samples of the Alacol set were collected during the period of rapid growth of vegetation. At the same time, the Shelek set was formed from samples collected on a plowed flat area with seedlings of low height, where salinity manifestations were expressed in relatively large areas of the field (hundreds of meters).

Large areas of the country and, as a consequence, the variability of the surrounding conditions leads to the fact that the field and satellite data from different areas may differ significantly, and their combined use requires the further research. Accounting for terrain parameters is one way to improve results. A possible solution is to apply deep learning models [63,77–80], which already demonstrate their advantages in some conditions [81].

It should be noted that UAVs [82] are actively used for solving the problems of precision agriculture, which make it possible to collect multispectral and hyperspectral data of high resolution, which are not available for satellite images [39,83,84]. The use of such data allows a detailed assessment of salinity within relatively small fields of a few hectares in size [85].

The results of mapping of some regions of South Kazakhstan and high-resolution illustrations are attached to the article as a Supplementary Materials.

6. Conclusions

Estimating soil salinity for practice, with sufficient accuracy and on the basis of remotely sensed data, is not an easy process. The main geophysical patterns of soil salinity are well understood, but soil salinity indicators are variable both spatially and temporally and depend significantly on the weather conditions, irrigation conditions and the moment of data collection in the course of field studies.

The experiments have shown that XGBRegressor has the best quality indicators among the considered regression models (R^2 = 0.654 for dataset Shelek). Although SAR data in general can be used to assess salinity in the southern regions of Kazakhstan, the performed computational experiments showed that the collected data sets are heterogeneous. On the wet soil (Shelek dataset), the results are significantly better, while the other two datasets separately show unsatisfactory results. It can be assumed that this is a consequence of the complexity of the landscape in the study area, the difference in the time of sampling, and the different nature and development of the vegetation cover. This fact requires additional analysis. Perhaps increasing the number of input parameters would help to improve the regression performance. A possible solution is to use those machine learning methods that are used for image processing, in other words, convolutional neural networks of various architectures.

The manifestation of salinity in optical and radar data is not the same. In order to assess the influence of input parameters, the article, probably for the first time, proposes a method for ranking and optimizing model input parameters using one of the EML methods—SHapley Additive exPlanations. In addition, in this work:

- A labeled data set is proposed for the electrical conductivity of soils in Southern Kazakhstan, which differ significantly in their geographical location;
- The method of soil salinity estimation described in [33] has been modified and extended with optical data;
- An analysis of several types of machine learning models was performed and it was shown that boosting regression models generally gives the best result;
- The results of the developed model are compared with the results of the Mtemp model [44] and it is shown that the developed model provides better agreement with ground-based measurements of electrical conductivity for this region.

Limitations of the study

1. This study is based on a relatively small amount of field data, which differ significantly in geophysical indicators of collection sites and collection times;
2. The quality of the work on the regressors significantly depends on the settings. Despite the search for the best combinations of parameters, it is not possible to analyze all combinations in a limited study;
3. The considered set of input parameters is not exhaustive. It is quite acceptable to use the remote sensing data both close to the time of sampling and remote in time.

Future research

In the future, it is planned to:

1. Evaluate the effect on the data of optical range, including infrared, on regression quality, depending on the time of remote sensing data acquisition;
2. Evaluate the impact of optical and radar data collected within the vegetation growth season (April–August) or for a longer period of time;
3. Apply deep learning models to account for terrain parameters;
4. Evaluate the possibilities of using multispectral images acquired from a UAV for mapping of focal salinity of agricultural fields.

Supplementary Materials: The following supporting information can be downloaded at: https://www.dropbox.com/sh/3i1e7ghg8r7kqwt/AAB3QjFMCKPD7ZDrC4fvPn5_a?dl=0.

Data Availability Statement: The data presented in this study are openly available in https://www.dropbox.com/sh/djlm3ox8briepv3/AAD691Fa2hPGtLIl5IEZ1Urka?dl=0, accessed on 1 August 2023.

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

Abbreviations

AI	Artificial intelligence
BC	Backscattering coefficient
BEX	Base exchange index
CARS	Competitive adaptive sampling with re-weighting
DEM	Digital elevation models
DSM	Digital surface model
EC	Electrical conductivity
ELV	Ground elevation model
EML	Explainable machine learning
FFNN	Feed-forward neural networks
GA	Genetic algorithm
GIS	Geographical information system
GLCM	Gray level co-occurrence matrix
GPS	Global positioning system
GQISWI	Groundwater quality index for seawater intrusion
LST	Land surface temperature
MAE	Mean absolute error
ML	Machine learning
MSE	Mean squared error
NDSI	Normalized differential salinity index
R^2	Coefficient of determination
RF	Random forest
RMSE	Root mean square error
SAR	Synthetic Aperture Radar
SHAP	SHapley Additive exPlanations
SRTM	Shuttle radar topography mission
SVM	Support vector machine
SWI	Seawater intrusion
UAVs	Unmanned aerial vehicles
varMAE	variance of MAE
varMSE	variance of MSE
VarR^2	variance of R^2
VIP	variable in projection

Appendix A Results of Machine Learning Models Using SAR Data

Dataset	Regressor	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Alakol	XGB	0.259	0.345	−1.166	0.027	0.354	2.621	15.78875
	RF	0.291	0.437	−6.881	0.026	0.307	135.626	18.61323
	LR	0.646	1.139	−39.321	0.048	0.857	3832.852	0.302191
	Lasso	0.257	0.386	−1.103	0.025	0.466	2.082	0.297237
	ElasticNet	0.257	0.386	−1.103	0.025	0.466	2.082	0.294212
	LGBM	0.257	0.386	−1.103	0.025	0.466	2.082	14.26181
	Ridge	0.322	0.468	−4.645	0.022	0.479	31.485	0.595438
	SVM	0.26	0.407	−1.787	0.028	0.466	12.747	0.342054
Kapchagay	XGB	0.695	3.47	−0.267	0.146	19.541	0.967	7.313395
	RF	1.132	5.466	−11.663	0.139	18.937	664.386	21.02584
	LR	1.412	5.04	−10.429	0.085	14.996	517.132	0.341087
	Lasso	0.821	3.57	−1.212	0.099	19.35	11.81	0.328122
	ElasticNet	0.821	3.57	−1.212	0.099	19.35	11.81	0.31415
	LGBM	0.772	3.501	−1.058	0.103	19.255	11.672	18.87905
	Ridge	0.905	3.798	−2.641	0.09	18.475	37.292	0.699129
	SVM	0.683	3.541	−0.34	0.138	20.373	1.377	0.373001
Shelek	XGB	0.665	0.864	**0.473**	0.032	0.209	0.038	50.68989
	RF	0.673	0.893	0.432	0.028	0.187	0.074	21.53441
	LR	0.813	1.25	0.186	0.037	0.567	0.245	0.352057
	Lasso	1.1	1.778	−0.104	0.028	0.374	0.02	0.330147
	ElasticNet	1.1	1.778	−0.104	0.028	0.374	0.02	0.311206
	LGBM	0.733	0.975	0.387	0.026	0.154	0.034	24.82316
	Ridge	0.821	1.016	0.349	0.019	0.125	0.038	0.698132
	SVM	0.888	1.243	0.233	0.031	0.296	0.04	0.386967
Full Dataset	XGB	0.644	1.991	**0.282**	0.024	3.542	0.046	28.14647
	RF	0.738	2.318	0.093	0.026	3.956	0.131	28.71924
	LR	0.913	2.515	−0.02	0.027	3.805	0.033	0.406411
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.380981
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.355016
	LGBM	0.814	2.178	0.183	0.033	3.721	0.029	11.27983
	Ridge	0.884	2.421	0.044	0.023	3.849	0.023	1.218743
	SVM	0.791	2.27	0.139	0.031	3.681	0.018	0.627324

Appendix B Results of Machine Learning Models Using SAR Data and Spectral Indices

Dataset	Regressor	MAE	MSE	R^2	VarMAE	VarMSE	VarR^2	Duration
Alakol	XGB	0.262	0.393	−0.713	0.035	0.419	2.186	22.36715
	RF	0.324	0.512	−7.397	0.034	0.392	201.885	19.18279
	LR	2.862	39.064	−1615.6	6.169	20,898.03	19,891,723	0.36007
	Lasso	0.289	0.47	−1.323	0.033	0.619	3.127	0.32513
	ElasticNet	0.289	0.47	−1.323	0.033	0.619	3.127	0.30421
	LGBM	0.289	0.47	−1.323	0.033	0.619	3.127	7.095027
	Ridge	0.379	0.56	−6.025	0.032	0.539	63.356	1.782234
	SVM	0.291	0.468	−1.398	0.032	0.608	3.588	0.345077
Kapchagay	XGB	0.66	3.153	−0.342	0.135	18.206	1.532	10.36428
	RF	0.9	4.667	−7.927	0.147	19.196	572.823	25.85497
	LR	–	–	–	–	–	–	–
	Lasso	0.821	3.57	−1.212	0.099	19.35	11.81	0.321146
	ElasticNet	0.821	3.57	−1.212	0.099	19.35	11.81	0.310198
	LGBM	0.757	3.335	−0.779	0.12	17.878	6.863	8.376564
	Ridge	0.919	4.044	−3.189	0.102	19.936	57.022	2.297855
	SVM	0.675	3.518	−0.286	0.137	20.212	1.221	0.424862
Shelek	XGB	0.508	0.579	**0.654**	0.02	0.101	0.018	46.30717
	RF	0.576	0.75	0.527	0.03	0.161	0.065	23.09414
	LR	–	–	–	–	–	–	–
	Lasso	1.1	1.778	−0.104	0.028	0.374	0.02	0.336103
	ElasticNet	1.1	1.778	−0.104	0.028	0.374	0.02	0.30322
	LGBM	0.655	0.793	0.508	0.024	0.115	0.024	9.146147
	Ridge	0.74	0.94	0.395	0.031	0.183	0.088	1.973743
	SVM	0.971	1.473	0.1	0.03	0.388	0.028	0.391953
Full Dataset	XGB	0.569	1.889	**0.339**	0.023	3.49	0.057	28.16568
	RF	0.692	2.349	0.058	0.032	3.925	0.276	39.8665
	LR	0.923	2.43	−0.095	0.023	2.712	0.184	0.409904
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.375001
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.338127
	LGBM	0.69	1.935	0.325	0.03	3.707	0.048	11.27485
	Ridge	0.795	2.211	0.145	0.02	3.66	0.055	4.66742
	SVM	0.811	2.242	0.145	0.029	3.573	0.015	0.791433

Appendix C Regressor Results Using an Optimized Set of Spectral Indices

Full Dataset	XGB	0.587	1.93	0.305	0.027	3.455	0.048	23.44931
	RF	0.706	2.485	−0.064	0.032	3.792	0.486	27.9018
	LR	0.8	2.147	0.136	0.021	3.152	0.057	0.404916
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.378987
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.352054
	LGBM	0.646	1.939	0.321	0.023	3.733	0.046	10.58872
	Ridge	0.732	2.043	0.266	0.02	3.742	0.037	1.028286
	SVM	0.556	1.958	**0.336**	0.032	3.936	0.059	0.628671

Appendix D Results of Regressors with Optimized SAR Dataset and Optical Indices

Full Dataset	XGB	0.575	1.858	**0.356**	0.023	3.526	0.061	36.6489
	RF	0.685	2.275	0.109	0.027	3.875	0.176	23.21074
	LR	0.88	2.412	−0.065	0.021	3.024	0.225	0.240527
	Lasso	0.991	2.522	−0.043	0.023	3.372	0.005	0.152631
	ElasticNet	0.991	2.522	−0.043	0.023	3.372	0.005	0.155417
	LGBM	0.671	1.971	0.314	0.023	3.89	0.051	3.863818
	Ridge	0.764	2.104	0.221	0.02	3.671	0.033	1.490418
	SVM	0.71	2.053	0.273	0.029	3.786	0.034	0.38091

N.B. The optimized set of input parameters includes: 'dem', 'temp', 'slope', 'dissimilarity_vv_1', 'contrast_vv_1', 'homogeneity_vv_1', 'correlation_vv_1', 'entropy_vv_1', 'dissimilarity_vh_1', 'contrast_vh_1', 'homogeneity_vh_1', 'correlation_vh_1', 'entropy_vh_1', 'gamma_vv_1', 'gamma_vh_1', 'NDSI1', 'S31', 'SI11', 'SI31', 'SI81', 'NDSIre1', 'SI3re1', 'SSRIre1', 'SSRI1'.

References

1. Li, X.; Wang, Z.; Song, K.; Zhang, B.; Liu, D.; Guo, Z. Assessment for salinized wasteland expansion and land use change using GIS and remote sensing in the west part of Northeast China. *Environ. Monit. Assess.* **2007**, *131*, 421–437.
2. Metternicht, G.I.; Zinck, J. Remote sensing of soil salinity: Potentials and constraints. *Remote Sens. Environ.* **2003**, *85*, 1–20.
3. Muhetaer, N.; Nurmemet, I.; Abulaiti, A.; Xiao, S.; Zhao, J. An Efficient Approach for Inverting the Soil Salinity in Keriya Oasis, Northwestern China, Based on the Optical-Radar Feature-Space Model. *Sensors* **2022**, *22*, 7226. [CrossRef]
4. Taghadosi, M.M.; Hasanlou, M.; Eftekhari, K. Soil salinity mapping using dual-polarized SAR Sentinel-1 imagery. *Int. J. Remote Sens.* **2019**, *40*, 237–252. [CrossRef]
5. Grissa, M.; Abdelfattah, R.; Mercier, G.; Zribi, M.; Chahbi, A.; Lili-Chabaane, Z. Empirical model for soil salinity mapping from SAR data. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 1099–1102.
6. Tripathi, A.; Tiwari, R.K. A simplified subsurface soil salinity estimation using synergy of SENTINEL-1 SAR and SENTINEL-2 multispectral satellite data, for early stages of wheat crop growth in Rupnagar, Punjab, India. *Land Degrad. Dev.* **2021**, *32*, 3905–3919. [CrossRef]
7. Mohamed, S.A.; Metwaly, M.M.; Metwalli, M.R.; AbdelRahman, M.A.; Badreldin, N. Integrating Active and Passive Remote Sensing Data for Mapping Soil Salinity Using Machine Learning and Feature Selection Approaches in Arid Regions. *Remote Sens.* **2023**, *15*, 1751.
8. Nurmemet, I.; Ghulam, A.; Tiyip, T.; Elkadiri, R.; Ding, J.-L.; Maimaitiyiming, M.; Abliz, A.; Sawut, M.; Zhang, F.; Abliz, A. Monitoring soil salinization in Keriya River Basin, Northwestern China using passive reflective and active microwave remote sensing data. *Remote Sens.* **2015**, *7*, 8803–8829. [CrossRef]
9. Singh, A.; Dwivedi, R. Delineation of salt-affected soils through digital analysis of Landsat MSS data. *Remote Sens.* **1989**, *10*, 83–92. [CrossRef]
10. Metternicht, G.; Zinck, J. Spatial discrimination of salt-and sodium-affected soil surfaces. *Int. J. Remote Sens.* **1997**, *18*, 2571–2586. [CrossRef]

11. Fernandez-Buces, N.; Siebe, C.; Cram, S.; Palacio, J. Mapping soil salinity using a combined spectral response index for bare soil and vegetation: A case study in the former lake Texcoco, Mexico. *J. Arid. Environ.* **2006**, *65*, 644–667. [CrossRef]

12. Masoud, A.; Koike, K. Arid land salinization detected by remotely-sensed landcover changes: A case study in the Siwa region, NW Egypt. *J. Arid. Environ.* **2006**, *66*, 151–167. [CrossRef]

13. Gabdullin, B.; Zhogolov, A.; Savin, I.Y.; Otarov, A.; Ibrayeva, M.; Golovanov, D. Application of multi-spectral satellite data for interpretation of soil salinization of the irrigated areas (case study of Southern Kazakhstan). *Vestn. Mosk. Univ. Seriya 5 Geogr.* **2016**, *5*, 34–41. Available online: https://vestnik5.geogr.msu.ru/jour/article/view/172/173 (accessed on 5 May 2023).

14. Gorji, T.; Yildirim, A.; Sertel, E.; Tanik, A. Remote sensing approaches and mapping methods for monitoring soil salinity under different climate regimes. *Int. J. Environ. Geoinform.* **2019**, *6*, 33–49.

15. Allbed, A.; Kumar, L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review. *Adv. Remote Sens.* **2013**, *2*, 373–385. [CrossRef]

16. Abbas, A.; Khan, S.; Hussain, N.; Hanjra, M.A.; Akbar, S. Characterizing soil salinity in irrigated agriculture using a remote sensing approach. *Phys. Chem. Earth Parts A/B/C* **2013**, *55*, 43–52.

17. Scudiero, E.; Skaggs, T.H.; Corwin, D.L. Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance. *Remote Sens. Environ.* **2015**, *169*, 335–343.

18. Rahmati, M.; Hamzehpour, N. Quantitative remote sensing of soil electrical conductivity using ETM+ and ground measured data. *Int. J. Remote Sens.* **2017**, *38*, 123–140. [CrossRef]

19. Fan, X.; Weng, Y.; Tao, J. Towards decadal soil salinity mapping using Landsat time series data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 32–41.

20. Qu, Y.-H.; Duan, X.-L.; Gao, H.-Y.; Chen, A.-P.; An, Y.-Q.; Song, J.-L.; Zhou, H.-M.; He, T. Quantitative retrieval of soil salinity using hyperspectral data in the region of Inner Mongolia Hetao irrigation district. *Spectrosc. Spectr. Anal.* **2009**, *29*, 1362–1366.

21. Dutkiewicz, A.; Lewis, M.; Ostendorf, B. Evaluation and comparison of hyperspectral imagery for mapping surface symptoms of dryland salinity. *Int. J. Remote Sens.* **2009**, *30*, 693–719.

22. Fallah Shamsi, S.R.; Zare, S.; Abtahi, S.A. Soil salinity characteristics using moderate resolution imaging spectroradiometer (MODIS) images and statistical analysis. *Arch. Agron. Soil Sci.* **2013**, *59*, 471–489.

23. Zhang, T.-T.; Qi, J.-G.; Gao, Y.; Ouyang, Z.-T.; Zeng, S.-L.; Zhao, B. Detecting soil salinity with MODIS time series VI data. *Ecol. Indic.* **2015**, *52*, 480–489.

24. Phonphan, W.; Tripathi, N.K.; Tipdecho, T.; Eiumnoh, A. Modelling electrical conductivity of soil from backscattering coefficient of microwave remotely sensed data using artificial neural network. *Geocarto Int.* **2014**, *29*, 842–859.

25. Zeng, W.; Zhang, D.; Fang, Y.; Wu, J.; Huang, J. Comparison of partial least square regression, support vector machine, and deep-learning techniques for estimating soil salinity from hyperspectral data. *J. Appl. Remote Sens.* **2018**, *12*, 022204.

26. Vermeulen, D.; van Niekerk, A. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. *Geoderma* **2017**, *299*, 1–12.

27. Akramkhanov, A.; Vlek, P.L. The assessment of spatial distribution of soil salinity risk using neural network. *Environ. Monit. Assess.* **2012**, *184*, 2475–2485. [PubMed]

28. Mukhamediev, R.I.; Symagulov, A.; Kuchin, Y.; Yakunin, K.; Yelis, M. From classical machine learning to deep neural networks: A simplified scientometric review. *Appl. Sci.* **2021**, *11*, 5541.

29. Allbed, A.; Kumar, L.; Aldakheel, Y.Y. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma* **2014**, *230*, 1–8.

30. Nosair, A.M.; Shams, M.Y.; AbouElmagd, L.M.; Hassanein, A.E.; Fryar, A.E.; Abu Salem, H.S. Predictive model for progressive salinization in a coastal aquifer using artificial intelligence and hydrogeochemical techniques: A case study of the Nile Delta aquifer, Egypt. *Environ. Sci. Pollut. Res.* **2022**, *29*, 9318–9340.

31. Mukhamediev, R.I.; Kuchin, Y.; Amirgaliyev, Y.; Yunicheva, N.; Muhamedijeva, E. Estimation of Filtration Properties of Host Rocks in Sandstone-Type Uranium Deposits Using Machine Learning Methods. *IEEE Access* **2022**, *10*, 18855–18872.

32. Wang, J.; Ding, J.; Yu, D.; Teng, D.; He, B.; Chen, X.; Ge, X.; Zhang, Z.; Wang, Y.; Yang, X. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci. Total Environ.* **2020**, *707*, 136092. [PubMed]

33. Hoa, P.V.; Giang, N.V.; Binh, N.A.; Hai, L.V.H.; Pham, T.-D.; Hasanlou, M.; Tien Bui, D. Soil salinity mapping using SAR Sentinel-1 data and advanced machine learning algorithms: A case study at Ben Tre Province of the Mekong River Delta (Vietnam). *Remote Sens.* **2019**, *11*, 128.

34. Merembayev, T.; Amirgaliyev, Y.; Saurov, S.; Wójcik, W. Soil Salinity Classification Using Machine Learning Algorithms and Radar Data in the Case from the South of Kazakhstan. *J. Ecol. Eng.* **2022**, *23*, 61–67.

35. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106.

36. Rivest, R.L. Learning decision lists. *Mach. Learn.* **1987**, *2*, 229–246.

37. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621.

38. Ma, G.; Ding, J.; Han, L.; Zhang, Z.; Ran, S. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. *Reg. Sustain.* **2021**, *2*, 177–188.

39. Yang, N.; Yang, S.; Cui, W.; Zhang, Z.; Zhang, J.; Chen, J.; Ma, Y.; Lao, C.; Song, Z.; Chen, Y. Effect of spring irrigation on soil salinity monitoring with UAV-borne multispectral sensor. *Int. J. Remote Sens.* **2021**, *42*, 8952–8978.

40. Wei, G.; Li, Y.; Zhang, Z.; Chen, Y.; Chen, J.; Yao, Z.; Lao, C.; Chen, H. Estimation of soil salt content by combining UAV-borne multispectral sensor and machine learning algorithms. *PeerJ* **2020**, *8*, e9087.

41. Guan, Y.; Grote, K.; Schott, J.; Leverett, K. Prediction of soil water content and electrical conductivity using random Forest methods with UAV multispectral and ground-coupled geophysical data. *Remote Sens.* **2022**, *14*, 1023.

42. Chen, B.; Zheng, H.; Luo, G.; Chen, C.; Bao, A.; Liu, T.; Chen, X. Adaptive estimation of multi-regional soil salinization using extreme gradient boosting with Bayesian TPE optimization. *Int. J. Remote Sens.* **2022**, *43*, 778–811.

43. Fathizad, H.; Ardakani, M.A.H.; Sodaiezadeh, H.; Kerry, R.; Taghizadeh-Mehrjardi, R. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma* **2020**, *365*, 114233.

44. Ivushkin, K.; Bartholomeus, H.; Bregt, A.K.; Pulatov, A.; Kempen, B.; de Sousa, L. Global mapping of soil salinity change. *Remote Sens. Environ.* **2019**, *231*, 111260.

45. Guan, X.; Wang, S.; Gao, Z.; Lv, Y. Dynamic prediction of soil salinization in an irrigation district based on the support vector machine. *Math. Comput. Model.* **2013**, *58*, 719–724. [CrossRef]

46. Wei, L.; Yuan, Z.; Yu, M.; Huang, C.; Cao, L. Estimation of arsenic content in soil based on laboratory and field reflectance spectroscopy. *Sensors* **2019**, *19*, 3904. [CrossRef]

47. Shahabi, M.; Jafarzadeh, A.A.; Neyshabouri, M.R.; Ghorbani, M.A.; Valizadeh Kamran, K. Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods. *Arch. Agron. Soil Sci.* **2017**, *63*, 151–160.

48. Khan, N.M.; Rastoskuev, V.V.; Shalina, E.V.; Sato, Y. Mapping salt-affected soils using remote sensing indicators—A simple approach with the use of GIS IDRISI. In Proceedings of the 22nd Asian Conference on Remote Sensing, Singapore, 5–9 November 2001.

49. Bannari, A.; Guedon, A.; El-Harti, A.; Cherkaoui, F.; El-Ghmari, A. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO-1) sensor. *Commun. Soil Sci. Plant Anal.* **2008**, *39*, 2795–2811. [CrossRef]

50. Tripathi, N.; Rai, B.K.; Dwivedi, P. Spatial modeling of soil alkalinity in GIS environment using IRS data. In Proceedings of the 18th Asian Conference in Remote Sensing, Kuala Lumpur, Malaysia, 20–24 October 1997; pp. A.8.1–A.8.6.

51. Nicolas, H.; Walter, C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* **2006**, *134*, 217–230.

52. Khan, N.M.; Rastoskuev, V.V.; Sato, Y.; Shiozawa, S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* **2005**, *77*, 96–109.

53. Abbas, A.; Khan, S. Using remote sensing techniques for appraisal of irrigated soil salinity. In Proceedings of the International Congress on Modelling and Simulation (MODSIM), Auckland, New Zealand, 10–13 December 2007; pp. 2632–2638.

54. Guo, B.; Zang, W.; Zhang, R. Soil salizanation information in the Yellow River Delta based on feature surface models using Landsat 8 OLI data. *IEEE Access* **2020**, *8*, 94394–94403. [CrossRef]

55. Yu, X.; Chang, C.; Song, J.; Zhuge, Y.; Wang, A. Precise monitoring of soil salinity in China's Yellow River Delta using UAV-borne multispectral imagery and a soil salinity retrieval index. *Sensors* **2022**, *22*, 546. [CrossRef]

56. USGS EROS Archive, Landsat Archives, Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) Level-1 Data Products. Available online: https://doi.org/10.5066/F71835S6 (accessed on 5 May 2023).

57. Richards, L.A. *Diagnosis and Improvement of Saline and Alkali Soils*; LWW: Philadelphia, PA, USA, 1954; Volume 78.

58. Measuring Soil Salinity. Available online: https://www.agric.wa.gov.au/soil-salinity/measuring-soil-salinity (accessed on 5 May 2023).

59. Scikit-Learn. Machine Learning in Python. Available online: https://scikit-learn.org/stable/ (accessed on 5 May 2023).

60. Pang, G.; Wang, T.; Liao, J.; Li, S. Quantitative Model Based on Field-Derived Spectral Characteristics to Estimate Soil Salinity in Minqin County, China. *Soil Sci. Soc. Am. J.* **2014**, *78*, 546–555. [CrossRef]

61. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. Explainable ai: A review of machine learning interpretability methods. *Entropy* **2020**, *23*, 18. [CrossRef] [PubMed]

62. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–10.

63. Mukhamediev, R.I.; Popova, Y.; Kuchin, Y.; Zaitseva, E.; Kalimoldayev, A.; Symagulov, A.; Levashenko, V.; Abdoldina, F.; Gopejenko, V.; Yakunin, K. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics* **2022**, *10*, 2552. [CrossRef]

64. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

65. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

66. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Paper ID 1786. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf (accessed on 1 August 2023).

67. Al Daoud, E. Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *Int. J. Comput. Inf. Eng.* **2019**, *13*, 6–10.

68. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]

69. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

70. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

71. Yu, H.-F.; Huang, F.-L.; Lin, C.-J. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach. Learn.* **2011**, *85*, 41–75. [CrossRef]

72. Santosa, F.; Symes, W.W. Linear inversion of band-limited reflection seismograms. *SIAM J. Sci. Stat. Comput.* **1986**, *7*, 1307–1330. [CrossRef]

73. Tikhonov, A.N.; Goncharsky, A.; Stepanov, V.V.e.; Yagola, A.G. *Numerical Methods for the Solution of Ill-Posed Problems*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1995; Volume 328.

74. Hoerl, A.E.; Kennard, R.W. Ridge regression: Applications to nonorthogonal problems. *Technometrics* **1970**, *12*, 69–82. [CrossRef]

75. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]

76. Li, J.; Zhang, T.; Shao, Y.; Ju, Z. Comparing machine learning algorithms for soil salinity mapping using topographic factors and Sentinel-1/2 data: A case study in the Yellow River delta of China. *Remote Sens.* **2023**, *15*, 2332. [CrossRef]

77. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

78. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

79. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708.

80. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015, Part III 18*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

81. Gu, Q.; Han, Y.; Xu, Y.; Ge, H.; Li, X. Extraction of saline soil distributions using different salinity indices and deep neural networks. *Remote Sens.* **2022**, *14*, 4647. [CrossRef]

82. Mukhamediev, R.I.; Symagulov, A.; Kuchin, Y.; Zaitseva, E.; Bekbotayeva, A.; Yakunin, K.; Assanov, I.; Levashenko, V.; Popova, Y.; Akzhalova, A. Review of some applications of unmanned aerial vehicles technology in the resource-rich country. *Appl. Sci.* **2021**, *11*, 10171. [CrossRef]

83. Wang, D.; Chen, H.; Wang, G.; Cong, J.; Wang, X.; Wei, X. Salinity inversion of severe saline soil in the yellow river estuary based on UAV multi-spectra. *Sci. Agric. Sin.* **2019**, *52*, 1698–1709.

84. Hu, J.; Peng, J.; Zhou, Y.; Xu, D.; Zhao, R.; Jiang, Q.; Fu, T.; Wang, F.; Shi, Z. Quantitative estimation of soil salinity using UAV-borne hyperspectral and satellite multispectral images. *Remote Sens.* **2019**, *11*, 736. [CrossRef]

85. Mukhamediev, R.; Amirgaliyev, Y.; Kuchin, Y.; Aubakirov, M.; Terekhov, A.; Merembayev, T.; Yelis, M.; Zaitseva, E.; Levashenko, V.; Popova, Y.; et al. Operational Mapping of Salinization Areas in Agricultural Fields Using Machine Learning Models Based on Low-Altitude Multispectral Images. *Drones* **2023**, *7*, 357. [CrossRef]