

Article

Not peer-reviewed version

Soil Salinity Inversion Based on a Stacking Integrated Learning Algorithm

[Haili Dong](#) and [Fei Tian](#) *

Posted Date: 27 August 2024

doi: [10.20944/preprints202408.1937.v1](https://doi.org/10.20944/preprints202408.1937.v1)

Keywords: soil salinity; machine learning; stacking; remote sensing inversion



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Soil Salinity Inversion Based on a Stacking Integrated Learning Algorithm

Haili Dong¹ and Fei Tian^{1,2,*}

¹ Center for Agricultural Water Research in China, China Agricultural University, Beijing 100083, China

² National Field Scientific Observation and Research Station on Efficient Water Use of Oasis Agriculture in Wuwei of Gansu Province, Wuwei 733000, China

* Correspondence: feitian@cau.edu.cn; Tel.: +86-010-6273-6273

Abstract: Soil salinization is an essential risk factor for agricultural development as well as for food security, and how to obtain regional soil salinity information more reliably remains a priority problem to be solved. To improve the accuracy of the inversions for soil salinity, a new inversion model for soil salinity based on stacking integrated algorithm for learning was submitted for this work which took the prediction results of several basic models as new features and then trained a secondary model to fuse the prediction results of basic models. We compared and analyzed it against four machine learning regression models, namely, random forest (RF), back propagation neural network, support vector regression, and convolutional neural network. Findings indicated the stacking integrated learning regression model fitted better and had good stability, on the test set, the stacking integrated learning regression model showed a relative increase of 8.16% in R^2 , a relative decrease of 13.95% in RMSE, and a relative increase of 6.47% in RPD when compared to the RF model, which was the single most effective machine learning regression model, and the stacking model was able to achieve soil salinity inversion more accurately. The soil salinity in the oasis areas of the Manas River Basin tended to decrease from north to south in 2016 to 2020 from a spatial point of view, and it was reduced in April from a temporal point of view. The percentage of pixels with a high soil salinity content of 2.75–2.8 g/kg in the study area had decreased by 19.64% in April 2020 compared to April 2016. The innovatively constructed stacking integrated learning regression model improved the accuracy of soil salinity estimation on the basis of the superior results obtained in the training of the single optimal machine learning regression model. As a consequence, this model can provide technological backup for a fast monitoring and inversion of soil salinity as well as prevention and containment of salinization.

Keywords: soil salinity; machine learning; stacking; remote sensing inversion

1. Introduction

Soil salinization is considered to be significant issues with ecological impacts, severely proscribing the safety and improvement of regional ecological areas [1–4]. Cultivated soil salinization is recognized to motive land degradation, harm crop growth, and preclude agricultural improvement [5–8]. Being in a position to gather soil salinity statistics rapidly and precisely ought to assist with efficaciously evaluating the diploma of soil salinization to facilitate the enhancement and utilization of salinized land. An ordinary technique for achieving soil salinity statistics has been fixed-point sampling with the use of a conductivity meter to measure the statistics [4,9–12]. Although this method has been shown to be effective, it has the shortcomings of being time consuming, being labor intensive, having poor representation of the measurement points, including only a small coverage area, and so forth and has limitations for soil salinity monitoring across large spatial scales [13]. In latest years, as remote sensing technology has been integrated with agriculture, it has been a means of achieving rapid acquisition of records at a lower value [4,14–16]. A growing number of remote sensing techniques are being applied to soil salinity monitoring [17–20]. Feature indices that are more sensitive to salinity are approached from remote sensing images, and a model is structured by incorporating feature indices with soil salinity content to facilitate the monitoring current status of

regional soil salinity [8]. This method can make up for the shortcomings of previous field surveys and allow researchers to study regional soil salinization from a larger spatial scale, with the advantages of obtaining information quickly, being less affected by the ground, and being able to continuously and dynamically monitor regional salinization status, making it one of the most widely used quantitative soil salinization monitoring methods today [21].

Inversion of empirical statistical regression models for biophysical parameters of vegetation using remote sensing is usually classified into simply linear regression models on the one hand and non-linear regression models on the other hand. In order to explore the feasibility of using multiple vegetation indexes to invert soil salinity content, Wu et al [22] estimated soil salinity primarily on the basis of linear regression, synergistic kriging, regression kriging, and geographically weighted regression, and the findings indicated that geographically weighted regression method was the most accurate (RMSE = 0.31), respectively. Although linear regression models could effectively reduce the uncertainty in the inversion process [23], these models are difficult to solve for data that are nonlinear or have high correlation between features. In order to accurately estimate soil water in semiarid areas in the western part of Khorasan-Razavi province in (northeast) Iran, Hamed et al. [24] explored the sensitivity of vegetation indexes calculated from Landsat 8 remote sensing imagery to soil water by using random forest (RF), elastic net regression, and linear regression models, and findings indicated that RF regression model was the most accurate (RMSE = 0.04). Nonlinear regression models can provide an explanation for the correlation between the bio-physical and model variables [25], are easy to parallelize, and have a relatively strong model generalization ability; however, it is often necessary to trade off the balance of such a model with its accuracy if the prediction error of a single regression model is relatively low [26]. Ghosh et al [27] estimated biomass of mangrove forests in India making use of series of machine learning algorithms just like RF model, gradient boosting model and extreme gradient boosting model as well as integration of multiple machine learning algorithms. They found that the accuracy of inverting the aboveground biomass was further improved, showing RMSE to be 72.87 t/ha, using a stacking algorithm in a multi-temporal image stacked dataset. This RMSE was 1.62 t/ha less than that of the single RF regression model, indicating that an integrated learning regression algorithm based on stacking could integrate multiple underlying regression models and provide enhanced generalization capabilities [28].

At present, the integrated learning regression model based on stacking has shown good performance in soil water inversion [29–31]; however, there is still a need for in-depth research in soil salinity content inversion. Therefore, we attempted to use the stacking integrated learning regression model in soil salinity inversion to obtain more accurate results. The study objectives are: (1) to establish an inverse soil salinity model which is mainly used for RF model, back propagation neural network (BPNN) model, SVR model, convolutional neural network (CNN) model, and integrated learning algorithms and (2) to evaluate and analyze accurately the established soil salinity model, to quantitatively characterize the characteristics of changes in the soil salinity content of the region, to provide accurate monitoring and an inversion model of soil salinity to provide technical support and a conceptual foundation for the future use of inversion models.

2. Materials and Methods

2.1. Study Area

The Manas River Basin Oasis Area lies in the mid-region of the northern foothills of the Tianshan Mountains in China, in the southern edge of the Junggar Basin with a longitude of 85°01'–86°32'E and a latitude of 43°27'–45°21'N, as shown in Figure 1. The area includes Lower Nodi Irrigation District, Anzhihai Irrigation District, Jinguhe Irrigation District, Shihezi Irrigation District, Xinhu General Field Irrigation District, and Mosuo Bay Irrigation District [32]. The basin is arid with an average annual temperature of 4.7°C–5.7°C, with the highest temperature in July and the lowest in January, an average annual precipitation of 100–200 mm, and, mainly concentrated in summer, an average annual evaporation of 1,500–2,100 mm [33,34]. Ice and snow meltwater in the region carry salts from rock weathering into farmland. This perennial salt aggregation has resulted in serious salinization of

the oasis, and due to the irrationality of irrigation, the salinization of farmland in the Manas River Basin is frequent, seriously affecting improvement of resource utilization and economic development [35]. Therefore, exploring fashions with greater accuracy is indispensable to understand regional soil salinity monitoring.

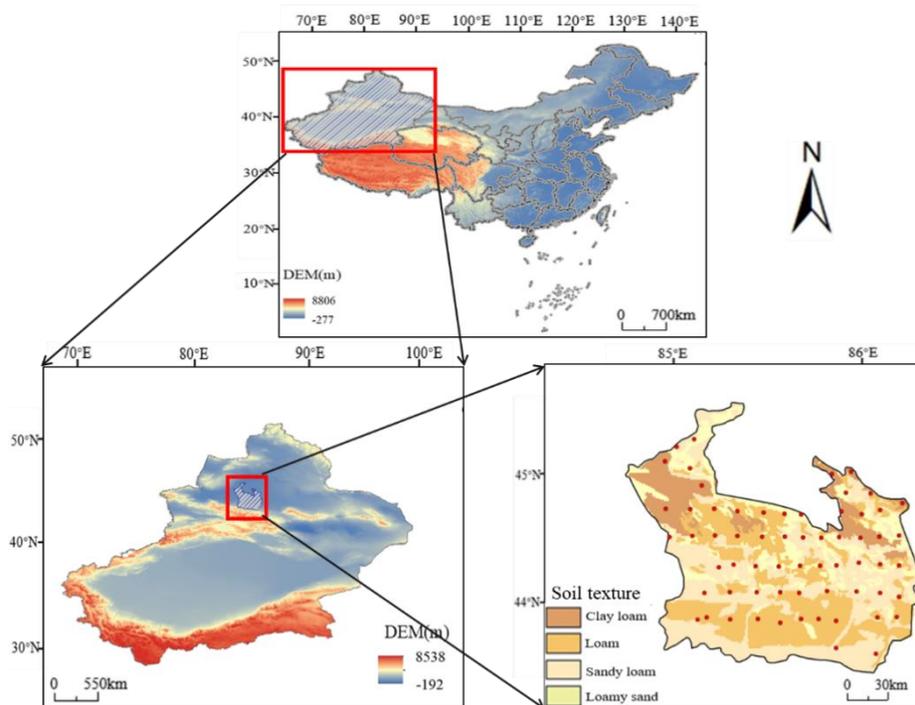


Figure 1. Distribution of oasis areas and sampling sites in the Manas River basin.

2.2. Data Sources

Soil salinity historical data (from April 2014) were sourced from reference 36 [36]. Remote sensing image data were synchronized with surface data, and the data were obtained from Landsat 8 OLI remote sensing image data collected by the United States Geological Survey (<http://glovis.usgs.gov/>). The data had been subjected in preprocessing steps such as geometric correction, radiometric correction, FLAASH atmospheric correction, synthesis, cropping, and so on. Moreover, visual interpretation and supervised classification had been applied to categorize the land use status to avoid any confusion regarding information about water bodies such as lakes, rivers, ponds, and puddles in the study area during salinity inversion.

2.3. Salinity Index Construction

The spectral index is an easy as well as valid method for measuring characteristic distributions on the land surface. As such, it was already widely used in global and regional land cover monitoring, vegetation classification and monitoring of environmental change [37–39]. In this study, according to the spectral characteristics of the features, we combined the spectral reflectance in various bands into the spectral index, and used them as indexes for remote sensing evaluation. Overall, five vegetation indexes, seven salinity indexes, one water index, and one brightness index were selected [5], and their formulas were shown in Table 1.

Table 1. Spectral index and their calculation formula.

Type of index	Spectral index	Abbrev	Formulas	Reference
---------------	----------------	--------	----------	-----------

Vegetation spectral indices (VI)	Normalized Difference Vegetation Index	NDVI	$\frac{NIR-R}{NIR+R}$	Shrestha et al. 2006[40]
	Difference Vegetation Index	DVI	$NIR-R$	Shrestha et al. 2006[40]
	Soil-Adjusted Vegetation Index	SAVI	$\frac{(1+L)(NIR-R)}{NIR+R+L} (L=0.5)$	Alhammadi et al. 2008[41]
	Ratio Vegetation Index	RVI	$\frac{NIR}{R}$	Alhammadi et al. 2008[41]
	Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR-G}{NIR+G}$	Bannari et al. 2018[15]
	Salinity Index	SI	$\sqrt{B*R}$	Yao Y et al. 2013[42]
	Salinity Index 1	SI1	$\sqrt{G*R}$	Allbed et al. 2014[2]
Salinity spectral indices (SI)	Salinity Index 2	SI2	$\sqrt{G^2+R^2+NIR^2}$	Douaoui et al. 2005[43]
	Salinity Index 3	SI3	$\sqrt{G^2+R^2}$	Douaoui et al. 2005[43]
	Salinity Index 7	SI7	$\frac{R*NIR}{G}$	Abbas et al. 2013[44]
	Normalized Difference Salinity Index	NDSI	$\frac{R-NIR}{R+NIR}$	Khan et al. 2001[45]
Soil Salinity Remote Sensing index	SRSI	$\sqrt{(NDVI-1)^2+SI1^2}$	Alhammadi et al. 2008[41]	
NDWI	Normalized Difference Water Index	NDWI	$\frac{G-NIR}{G+NIR}$	Liu H J et al. 2018[46]
BI	Brightness Index	BI	$\sqrt{R^2+NIR^2}$	Khan et al. 2001[45]

Note: NIR, R, G and B are the reflectance in the near-infrared, red, green and blue light bands of the Landsat8-OIL satellite, respectively; L is the soil conditioning factor, which is generally taken as 0.5 to reduce the effect of reflections from the soil surface.

2.4. Model Construction and Accuracy Evaluation

2.4.1. Model Construction and Model Parameters Determination

The input and output data of the soil salinity inversion model were soil spectral index values calculated by Landsat 8 and surveyed soil salinity data, respectively, and model were built by using RF, BPNN, SVR and CNN models.

The grid search method [47] was used to identify the best variables in the machine learning regression model. The number of decision trees of the RF model was set to 200, and the minimum samples required for the division of internal nodes was set to 5. The number of iterations of the BPNN model was set to 1,500, the error threshold was set to 1e-5, and the learning rate was set to 0.001. The number of iterations of the CNN model was set to 1,000, the error threshold was set to 1e-6, and the learning rate was set to 0.01. Finally, the kernel parameter type of the SVR model was set to 0.01, the kernel function of the SVR model was set to "linear", and the box constraint was set to 1.

The main steps of the stacking integrated learning regression model were as follows: (1) use MATLAB logic statements to optimize the results of the model which has the best model precision in a single machine learning regression algorithm, (2) use the trained RF model and BPNN model to predict a training set and a test set, (3) link the training set and the test set to the predicted results of

RF and BPNN model to form a new training set and test set, (4) construct an stacking integrated learning regression model, and (5) use the trained integrated model to predict a training set and test set to evaluate performance of the model.

There are structure and rationale of the stacking integrated learning regression model as shown in Figure 2. The technology roadmap is shown in Figure 3.

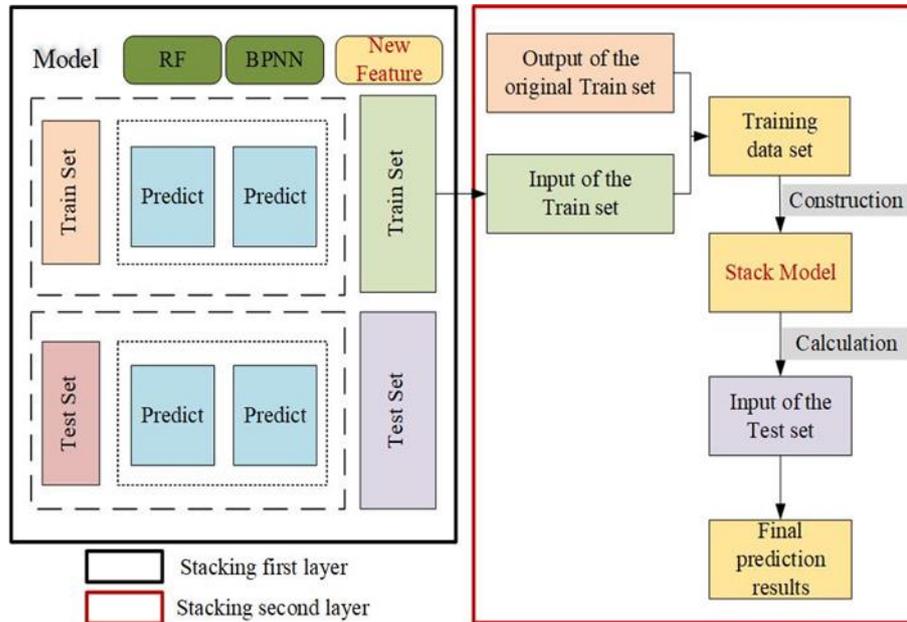


Figure 2. Principle of the Stacking integrated learning regression model.

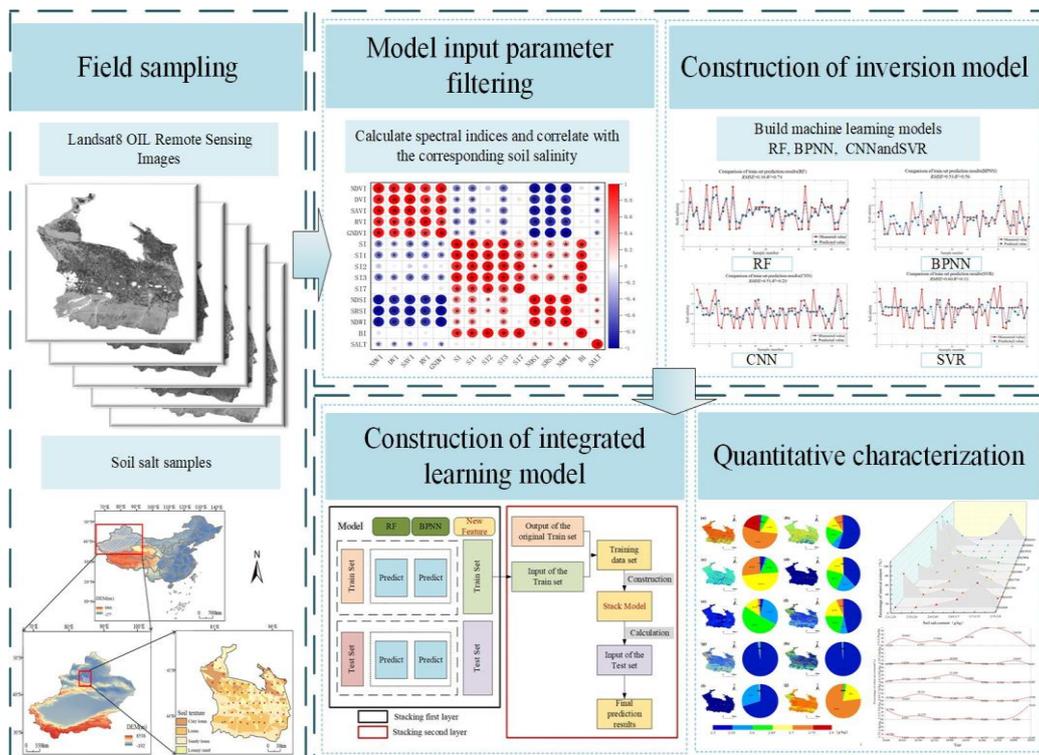


Figure 3. Research technology roadmap.2.4.2 Selection of model performance indicators.

The accuracy of the model that was assessed by the coefficient of determination (R^2), RMSE, and relative percentage difference (RPD) [4,8]. Generally, higher R^2 values and smaller RMSE values indicate a more favourable model. The RPD was classified in three levels, which were Class A (RPD

> 2.0, the constructed model is considered to be highly reliable), Class B ($1.40 \leq RPD \leq 2.0$, the constructed model is considered to be moderately reliable), and Class C ($RPD < 1.40$, the constructed model is not reliable) [4,5,36]. The formula to calculate the evaluation indicators is as follows:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (1)$$

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

$$RPD = \frac{\sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n}}}{RMSE} \quad (3)$$

where \hat{y}_i , y_i , and \bar{y}_i are the predicted, measured, and average measured values of soil salinity, respectively, and n is the number of samples.

3. Results

3.1. Correlation Analysis between Spectral Indexes and Soil Salinity

There were 14 soil spectral indicators and soil salinity derived from Origin software, which conformed to the normal distribution test, the results of Pearson correlation analysis are shown in Figure 4.

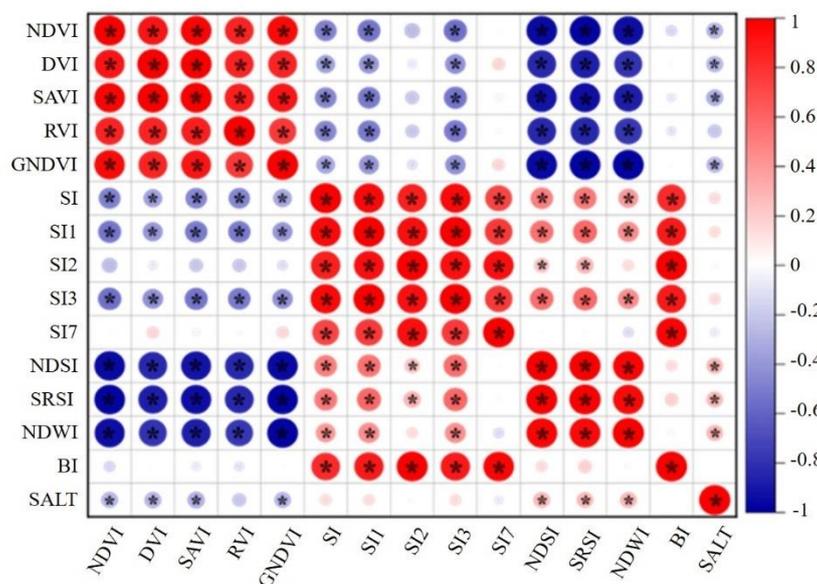


Figure 4. Correlation between spectral index and soil salinity. *: Correlation was significant at the 0.01 level (two-tailed).

As seen in Figure 4, among the triangular vegetation index, salinity index, water index, and brightness index, only the brightness index has no significant correlation with soil salinity, indicating that the relationship between soil salinity and the brightness index is relatively weak. Normalized Difference Snow Index (NDSI) and Standardized Reservoir Supply Index (SRSI) are positively correlated with soil salinity, and these indexes have large values in high salinity areas, reflecting the accumulation of soil salinity. Normalized Difference Water Index (NDWI) also shows a positive correlation with soil salinity. The Normalized Difference Vegetation Index (NDVI), Difference Vegetation Index (DVI), Soil-Adjusted Vegetation Index (SAVI), and Green Normalized Difference Vegetation Index (GNDVI) are negatively correlated with soil salinity, which indicates that in areas

of high salinity, the growth and cover of vegetation was poorer and high salinity inhibited the normal growth of plants, resulting in a negative correlation.

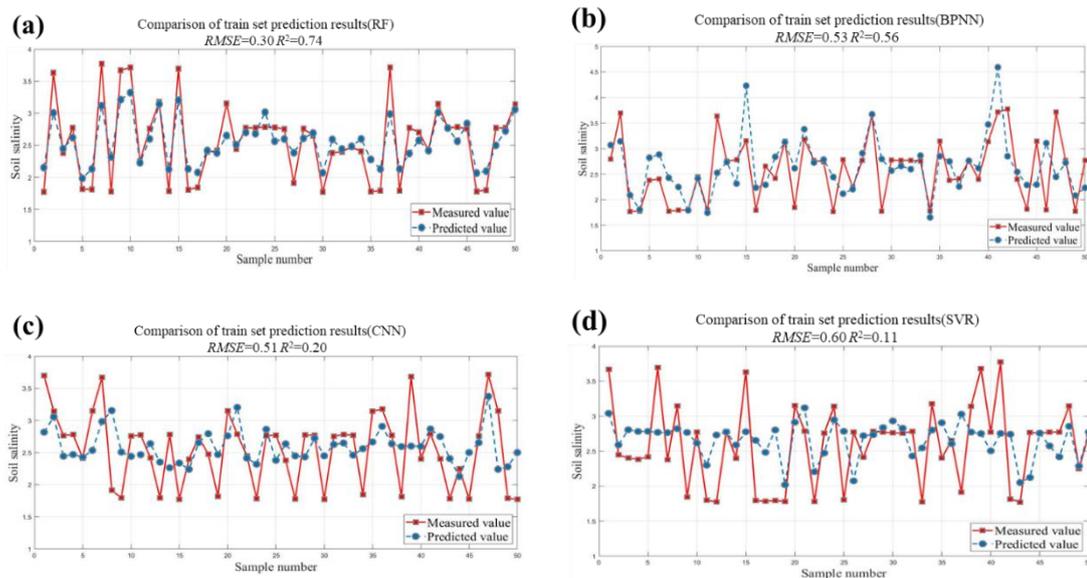
3.2. Evaluation of Machine Learning Regression Models

The spectral index data (NDSI, SRSI, NDWI, NDVI, DVI, SAVI, and GNDVI) with good correlation and the corresponding soil salinity data were used as the input data of models. Four machine learning methods (RF, BPNN, CNN, and SVR) were used to construct the soil salinity inversion models, prediction results are shown in Table 2. Forecast values of four inversion models of soil salinity had been in comparison against the measurements of soil salinity, as shown in Figure 4.

Table 2. Inversion model results of soil salinity based on spectral indices.

Model	Train Set			Test Set		
	R^2	RMSE	RPD	R^2	RMSE	RPD
RF	0.74	0.30	1.98	0.49	0.43	1.39
BPNN	0.56	0.53	1.51	0.26	0.52	1.21
CNN	0.20	0.51	1.13	0.18	0.54	1.08
SVR	0.11	0.60	1.07	0.02	0.62	1.06

As seen in Table 2 and Figure 5, the single optimal machine learning regression model was the RF. On the training set, both the RF and BPNN achieved better inversion results with R^2 above 0.5, RMSEs of 0.30 and 0.53, respectively, and RPDs of 1.4 or more, which reflects high accuracy. meanwhile, the CNN and SVR models underperform poorly on the training set, with R^2 of 0.20 and 0.11, RMSEs of 0.51 and 0.60, and RPDs below 1.4, respectively, indicating that the constructed models were unreliable. On the test set, the best inversion model was RF which had R^2 of 0.49, RMSE of 0.43, and RPD of 1.40, with accuracies 23%, 31%, and 47% higher than those of the BPNN, CNN, and SVR models, respectively; the R^2 of the four models, in descending order, was RF > BPNN > CNN > SVR, and the RMSEs were the opposite of this. According to the model performance evaluation index, it was reasonable to use the spectral index to establish the model. Therefore, in this paper, the RF and BPNN models with better inversion results were selected as the base models for the stacking integrated learning regression.



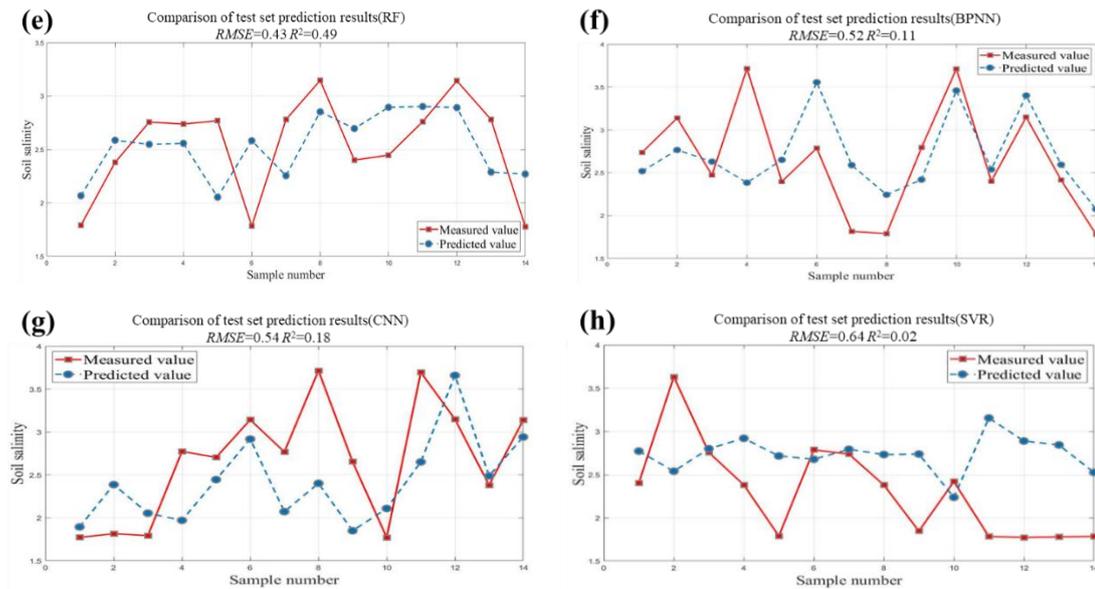


Figure 5. Prediction results for RF, BPNN, CNN and SVR (a, b, c and d are the prediction results for the RF, BPNN, CNN and SVR training set, and e, f, g and h are the prediction results for the RF, BPNN, CNN and SVR test set, respectively).3.3. Stacking integrated learning regression model evaluation.

The stacking integrated learning regression model can integrate many basic regression models and so forth to create much stronger predictions and deliver enhanced capabilities of generalization during inversion. The prediction results of the training set and test set of stacking integrated learning regression model are shown in Figure 6.

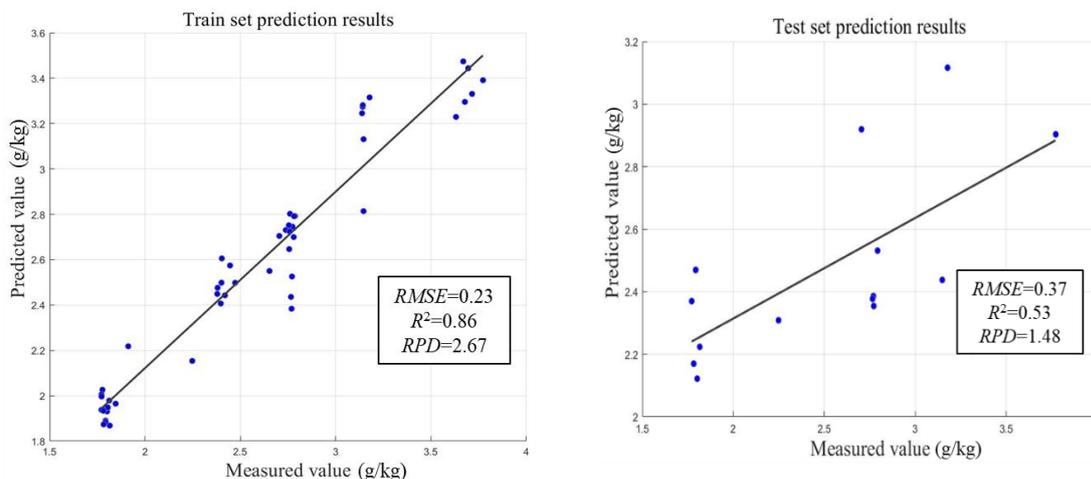


Figure 6. Scatter plot of measured and predicted values of Stacking integrated learning regression model. As seen in Figure 6, the stacking integrated learning regression model has high integration capability, and the model fit is strong as derived from the fitting curves. On the training set, the R^2 was 0.86, the RMSE was 0.23, and the RPD was 2.67, and on the test set, the R^2 was 0.53, the RMSE was 0.37, and the RPD was 1.48, which shows that the model has good stability.

A performance comparison of the results of stacking integrated learning, RF, BPNN, CNN, and SVR regression models is shown in Figure 7. As seen in Figure 6, on the training set, the R^2 of the stacking integrated learning regression model relatively improved by 16.22%, the RMSE relatively reduced by 23.33%, the RPD relatively improved by 34.85% compared with the single optimal RF model, which changed from being a moderately reliable model to a model with higher reliability that

can be used for model analysis. On the test set, the R^2 of the stacking integrated learning regression model was relatively improved by 8.16%, the RMSE was relatively reduced by 13.95%, and the RPD was relatively improved by 6.47% compared with the RF. In summary, the integrated learning regression model had a stronger generalization ability and improved the estimation accuracy of soil salinity once again on the basis of the superior results obtained in the training of the single optimal machine learning regression model.

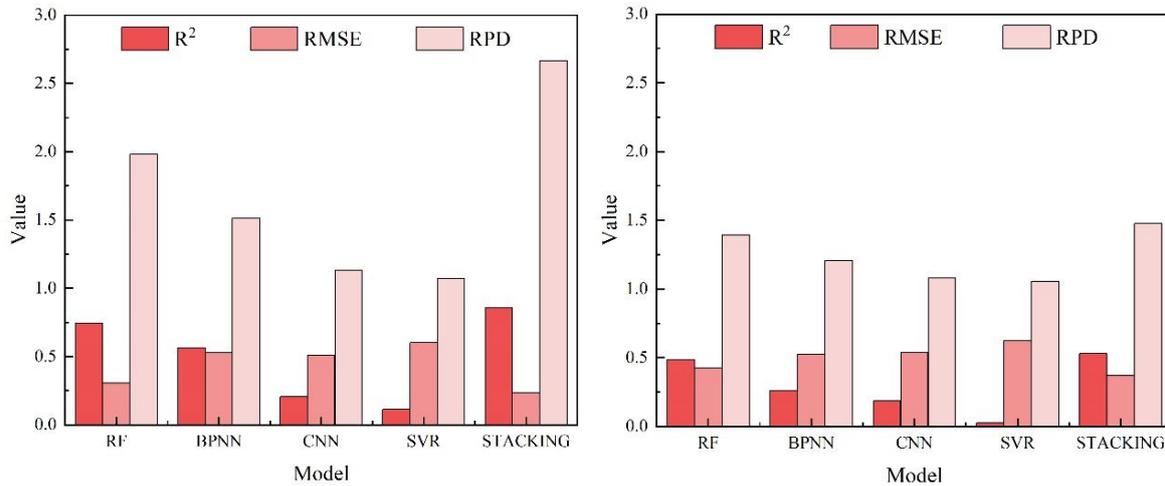
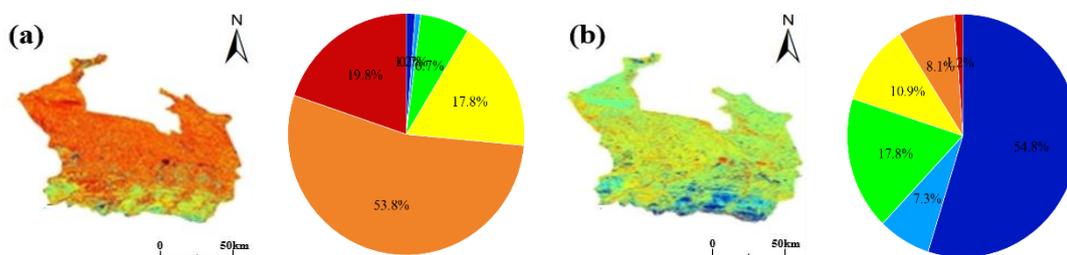


Figure 7. Performance evaluation of Stacking integrated learning, RF, BPNN, CNN and SVR regression models. Note: The left figure is the training set and the right figure is the test set.

3.4. Spatiotemporal Distribution of Soil Salinity

To better reflect inversion effect of model, the spatial distribution of soil salinity in study area in April and October 2016–2020 was mapped based on the constructed stacking integrated learning regression model. As seen in Figure 8, there was a significant difference in soil salinity. From a spatial point of view, the soil salinity content in study area in April and October of 2016, 2017, and 2018 showed a decreasing trend from north to south, with more pronounced salinization in the east-central part (near the Manas power plant and the town of Anjihai). In 2016, 2017, and 2018, the average salt content in April was 2.71, 2.67, and 2.61 g/kg, respectively, and the average salt content in October was 2.53, 2.57, and 2.58 g/kg, respectively. The values from October were lower by 0.18, 0.10, and 0.03 g/kg, respectively, compared to April. There was no significant change in soil salinity in most areas in April and October 2019. The average value of soil salinity in October 2020 was 2.71 g/kg, which was significantly higher, by 7.11%, compared to that of April of the same year.



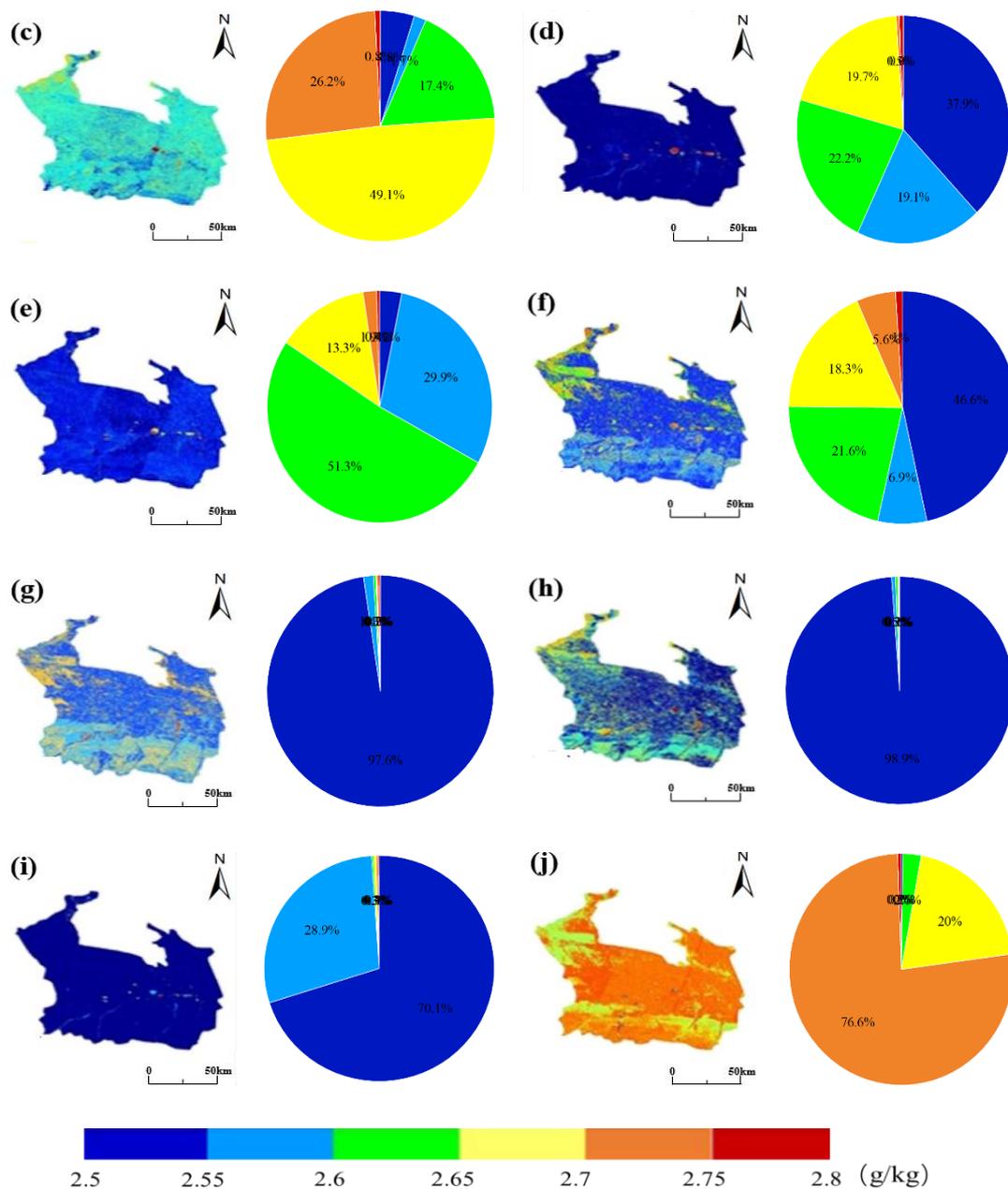


Figure 8. Spatial distribution of soil salinity in April and October 2016-2020 (a, b, c, d and e were for April 2016-2020; f, g, h, i and j were for October 2016-2020). Note: Pie charts were regional shares of salinity content.

As can be seen from Figure 9, the percentage of pixels with high soil salinity content of 2.75–2.8 g/kg in the study area in April 2016, October 2016, April 2020, and October 2020 was 19.78%, 1.16%, 0.14%, and 0.51%, respectively. These values decreased by 19.64% in April 2020 compared to April 2016, indicating a reduction in salinization. From 2016 to 2018, 55.30% of the area's soil salinity shifted from 2.65–2.80 g/kg to 2.50–2.65 g/kg, especially from October to April. Soil salinity was maintained at 2.50–2.55 g/kg in 70.07% of the area in 2019. However, soil salinity intensified in October compared to April in 2020, with 96.34% of the area shifting from 2.55–2.65 g/kg to 2.65–2.80 g/kg, probably due to the disturbance of vegetation cover when acquiring remote sensing images. Overall, in April, the degree of soil salinization in the oasis area of the Manas River Basin decreased gradually, and the saline-alkaline land improvement measures were still effective.

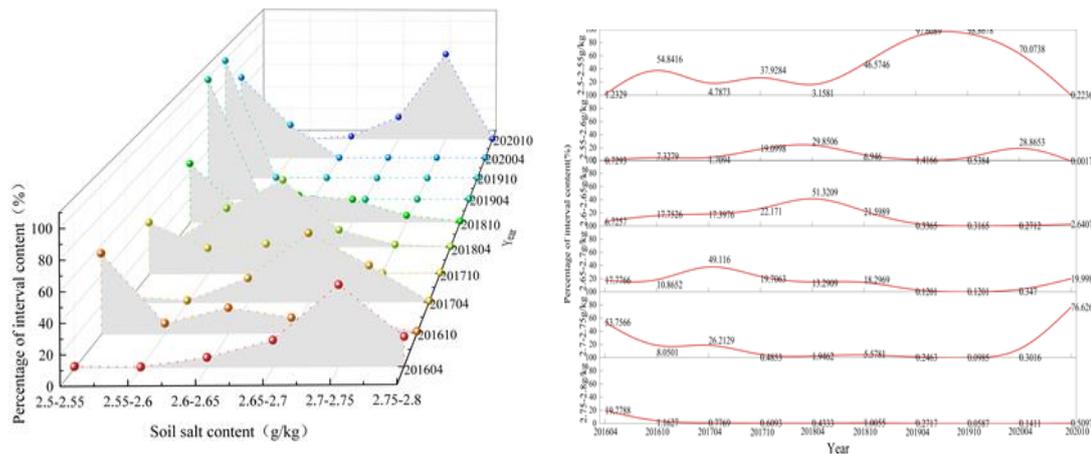


Figure 9. Soil salinity content over time.

4. Discussion

Accurately monitoring soil salinity content is essential to both food production with precision agriculture construction [48]. In this paper, a Pearson correlation analysis was introduced to filter the characteristics in spectral indices [49], and the findings of this study indicated that dominant factors of soil salinity were mainly vegetation index, salinity index, and water index, and this is in agreement to Alhammadi et al. [41–44]. The difference was that this study obtained a significant correlation between water index and soil salinity. Soil salinity migrates when soil moisture moves, and there will be an increase in soil electrical conductivity and higher soil moisture content in areas with higher soil salinity [50,51]. Tang et al [52] also find that the albedo of salinized soil increases due to the specular effect of water bodies as soil moisture content increases towards a critical point. Therefore, the water index, vegetation index, and salinity index were input as variables for the building of soil salinity inversion models.

The development of machine learning algorithms has accelerated the remote sensing modeling process. Many machine learning algorithms often exceed the accuracy of traditional regression modeling [40–42]. In this research, four machine learning regression models, RF, BPNN, CNN and SVR, had been made to simulate soil salinity, of which RF obtained best inversion results with R^2 of 0.74, RMSE of 0.30 and RPD of 1.98. Wei et al. [53] quantitatively estimated soil salinity by using multispectral imagery from unmanned aerial vehicles and established the BPNN, SVR, and RF models, and their findings indicated RF model was the most effective. Zhang et al. [54] explore the problem of rapid monitoring of soil salinity. They used support vector machine, BPNN, RF, and multivariate linear regression models to establish soil salinity, and their findings indicated RF model is the optimal with an R^2 of 0.72, this is in agreement well with this research.

Because of both plant individual differences and changes in canopy over the seasons, there is often the shortcoming of a single machine learning model with poor generalizations [8]. Yang et al [8,55] presented a stacking approach to accomplish the task of efficiently estimating highly accurate forecasts by combining a number of weak forecasters into one powerful forecaster, which can effectively improve the estimation accuracy. Presently, while integrated learning Models on the basis of stacking algorithms are broadly applied for the domains of machine vision and natural language processing, the application of integrated learning models for soil salinity have not been explored in terms of stacking strategy [8]. In this study, we constructed a stacking-based integrated regression model that produces higher accuracy, with an R^2 increase more than 16.22% compared to a single machine learning regression model based on a training set. Thus, there is a significant advantage of the stacking algorithm in predicting soil salinity. In other words, the more complex integration algorithm, that is, the stacking integration model, is significantly more capable of handling complex problems than the single machine learning regression model, a finding also supported by Pham et al. [56]. The first layer of the stacking method consists with various underlying models with inputs from

the original training set, while the outputs are the predicted values of the various base models. The second layer consists of only one meta-model, which is trained on forecast and true values of the various models in the first layer to form the completely integrated model. Similarly, the process of predicting the test set goes through the predictions of all the base models to compose the features of the second layer, next, a second layer of meta-model is implemented to predict final outputs in order to keep approaching the true value [57], which could be prone to overfitting if the training set of the base models is used directly to generate the training set of the meta-model. This study focused on RF and BPNN based models rather than those based on the CNN and SVR. This was because the RF and BPNN based models obtained better results than the CNN and SVR based models when we tested the model performance. However, the approach to combining the advantages of a single model to build an integrated model remains the key to future research.

The stacking integrated learning regression model was constructed to carry out soil salinity inversion with the oasis zone of the Manas River Basin as the research object. Findings of research show soil salinity content in this area had a decreasing trend from north to south in 2016, 2017, and 2018. Region of study is a typical mountain-basin structural system and sees differences in soils, water table, and climate based on latitude and location [58]. Zhang et al. found that a low-latitude area had a long development time, high soil maturity, and low soil salinity. Meanwhile, a mid-latitude area had a high water table, and with the promotion of membrane drip irrigation technology, the crop area could expand rapidly. Moreover, the membrane drip irrigation technology did not take away the salinity, so the mid-latitude area had a higher soil salinity. Finally, a high-latitude area that situated at the lower river basin and had a texture that was dominated by sandy loam and sandy soil would have a low water table. The soil salinity would also be higher than that in the mid-latitude area [59–61], this is in agreement with this study's results. In 2016–2018, soil salinity was reduced in October compared with April, probably because the change in groundwater burial depth during the year showed a mining type. Then in October, the groundwater declined rapidly, and the deeper groundwater burial depth and frequent agricultural irrigation made soil salinity decrease in the month. It is in accordance with the findings of Chen et al. [62] on soil salinity and nutrients in different landscape types. However, in April and October 2019, there was no significant change in soil salinity in the vast majority of the area. Soil salinity became stronger in 2020 compared to 2019, especially in October 2020, when soil salinity increased significantly. As temperatures increased, soil water rose vigorously, and salts then increased and accumulated in the soil surface. This has been similar to Zhao et al.'s [63] study on 121 Corps, 8th Agricultural Division, Manas River Basin, where crop soil salinity showed a salt accumulation trend in early May and mid-September. The integrated learning regression model constructed in this study is a fast and accurate inversion of soil salinity content for quantitative assessment and monitoring of saline soils regionally. Nevertheless, the distribution patterns of soil salinity for various crops and seasons may have significant differences that would need to be further investigated in the future.

5. Conclusions

We constructed soil salinity inversion models by combining Landsat 8 OLI remote sensing data and four methods of machine learning. Findings indicated that the single optimal machine learning regression model was the RF. In addition, the innovative stacking integrated learning regression model that was constructed in this study yielded better soil salinity inversion results, with a relative increase of 16.22% in the R^2 , a relative decrease of 23.33% in the RMSE, and a relative increase of 34.85% in the RPD on the training set compared with the single optimal machine learning regression model. On the test set, the R^2 was relatively improved by 8.16%, the RMSE was relatively reduced by 13.95%, and the RPD was relatively improved by 6.47%. Soil salinity in the oasis area of the Manas River Basin was accurately predicted and characterized. The soil salinity content in this area in April and October of 2016, 2017, and 2018 shows a trend of decreasing from north to south. And it gradually decreased in April, with the percentage of pixels with high soil salinity content of 2.75–2.8 g/kg in the study area decreasing by 19.64% in April 2020 compared to April 2016. Overall, the regression model on the basis of stacking had high accuracy, which could be of important relevance towards the quick

abstraction of salinity status and soil salinity maintenance and management in the Manas River Basin. However, further exploration is needed to combine advantages of multiple single models to construct an integrated model, and this consideration is the key to future research.

Author Contributions: Conceptualization, H.D. and F.T.; Methodology, H.D. and F.T.; Software, H.D.; Validation, H.D. and F.T.; Formal Analysis, H.D.; Investigation, H.D.; Resources, F.T.; Data Curation, H.D.; Writing – Original Draft Preparation, H.D.; Writing – Review & Editing, H.D. and F.T.; Visualization, H.D.; Supervision, F.T.; Project Administration, F.T.; Funding Acquisition, F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China (2022YFD190050403), and the National Key Research and Development Program of China (2021YFD1900800), and Western Light-Key Laboratory Cooperative Research Cross-Team Project of Chinese Academy of Sciences(xbzg-zdsys-202103).

Data Availability Statement: Not applicable.

Conflicts of Interest: All the authors declare no conflict of interest.

References

1. Ivushkin K, Bartholomeus H, Bregt A K, et al. Global mapping of soil salinity change[J]. *Remote Sensing of Environment*, 2019, 231: 111260.
2. 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[J]. *Geoderma*, 2014, 230-231: 1-8.
3. Metternicht G I, Zinck J A. Remote sensing of soil salinity: potentials and constraints[J]. *Remote Sensing of Environment*, 2003, 85(1): 1-20.
4. Zhao W, Zhou C, Zhou C, et al. Soil Salinity Inversion Model of Oasis in Arid Area Based on UAV Multispectral Remote Sensing[J]. *Remote Sensing*, 2022, 14(8): 1804.
5. Peng J, Biswas A, Jiang Q, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. *Geoderma*, 2019, 337: 1309-1319.
6. Wang Z, Zhang X, Zhang F, et al. Estimation of soil salt content using machine learning techniques based on remote-sensing fractional derivatives, a case study in the Ebinur Lake Wetland National Nature Reserve, Northwest China[J]. *Ecological Indicators*, 2020, 119: 106869.
7. Wang S, Chen Y, Wang M, et al. SPA-Based Methods for the Quantitative Estimation of the Soil Salt Content in Saline-Alkali Land from Field Spectroscopy Data: A Case Study from the Yellow River Irrigation Regions[J]. *Remote Sensing*, 2019, 11(8): 967.
8. Ding Y, Zhang J. Estimation of SPAD value in tomato leaves by multispectral images[J]. *Journal of Physics: Conference Series*, 2020, 1634(1):012128 (7pp).
9. Narjary B ,Meena, M.D, Kumar S ,et al. Digital mapping of soil salinity at various depths using an EM38[J].*Soil Use and Management*, 2018.
10. Csillag F, Pásztor L, Biehl L L. Spectral band selection for the characterization of salinity status of soils[J]. *Remote Sensing of Environment*, 1993, 43(3): 231-242.
11. Eldeiry A A, Garcia L A. Detecting Soil Salinity in Alfalfa Fields using Spatial Modeling and Remote Sensing[J]. *Soil Science Society of America Journal*, 2008, 72(1): 201-211.
12. Kalra N K, Joshi D C. Potentiality of Landsat, SPOT and IRS satellite imagery, for recognition of salt affected soils in Indian Arid Zone[J]. *International Journal of Remote Sensing*, 1996, 17(15): 3001-3014.
13. Ding J, Yu D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan–Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. *Geoderma*, 2014, 235-236: 316-322.
14. Ma Y, Chen H, Zhao G, et al. Spectral Index Fusion for Salinized Soil Salinity Inversion Using Sentinel-2A and UAV Images in a Coastal Area[J]. *IEEE Access*, 2020, 8: 159595-159608.
15. Bannari A, El-Battay A, Bannari R, et al. Sentinel-MSI VNIR and SWIR Bands Sensitivity Analysis for Soil Salinity Discrimination in an Arid Landscape[J]. *Remote Sensing*, 2018, 10(6): 855.
16. Dehni A, Lounis M. Remote Sensing Techniques for Salt Affected Soil Mapping: Application to the Oran Region of Algeria[J]. *Procedia Engineering*, 2012, 33: 188-198.
17. Aldabaa A A A, Weindorf D C, Chakraborty S, et al. Combination of proximal and remote sensing methods for rapid soil salinity quantification[J]. *Geoderma*, 2015, 239-240: 34-46.
18. Li Y, Wang C, Wright A, et al. Combination of GF-2 high spatial resolution imagery and land surface factors for predicting soil salinity of muddy coasts[J]. *CATENA*, 2021, 202: 105304.
19. Xu Y, Smith S E, Grunwald S, et al. Estimating soil total nitrogen in smallholder farm settings using remote sensing spectral indices and regression kriging[J]. *CATENA*, 2018, 163: 111-122.

20. An D, Zhao G, Chang C, et al. Hyperspectral field estimation and remote-sensing inversion of salt content in coastal saline soils of the Yellow River Delta[J]. *International Journal of Remote Sensing*, 2016, 37(2): 455-470.
21. Qiu Y, Chen C, Han J, et al. Satellite remote sensing estimation modeling of soil salinity in the irrigation domain of Jiefangzha under vegetation cover conditions[J]. *Water Saving Irrigation*, 2019(10): 108-112. CNKI:SUN:JSGU.0.2019-10-024.
22. Wu C, Liu G, Huang C. Prediction of soil salinity in the Yellow River Delta using geographically weighted regression[J]. *Archives of Agronomy and Soil Science*, 2017, 63(7): 928-941.
23. Lin C Y, Lin C. Using Ridge Regression Method to Reduce Estimation Uncertainty in Chlorophyll Models Based on Worldview Multispectral Data[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. 2019: 1777-1780.
24. Adab H, Morbidelli R, Saltalippi C, et al. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data[J]. *Water*, 2020, 12(11): 3223.
25. Tang S F, Tian Q J, Xu K J, et al. Inversion of larch forest age information by Sentinel-2 satellite[J]. *Journal of Remote Sensing*, 2020, 24(12): 1511-1524.
26. Christensen S W. Ensemble Construction via Designed Output Distortion[M]//Windeatt T, Roli F. *Multiple Classifier Systems*: 2709. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003: 286-295.
27. Ghosh S M, Behera M D, Jagadish B, et al. A novel approach for estimation of aboveground biomass of a carbon-rich mangrove site in India[J]. *Journal of Environmental Management*, 2021, 292: 112816.
28. Dietterich T G. Ensemble Methods in Machine Learning[C]//Multiple Classifier Systems. Berlin, Heidelberg: Springer, 2000: 1-15.
29. Tao S, Zhang X, Feng R, et al. Retrieving soil moisture from grape growing areas using multi-feature and stacking-based ensemble learning modeling[J]. *Computers and Electronics in Agriculture*, 2023, 204: 107537.
30. Qi J, Zhang X, McCarty G W, et al. Assessing the performance of a physically-based soil moisture module integrated within the Soil and Water Assessment Tool[J]. *Environmental Modelling & Software*, 2018, 109: 329-341.
31. Wang S, Wu Y, Li R, et al. Remote sensing-based retrieval of soil moisture content using stacking ensemble learning models[J]. *Land Degradation & Development*, 2023, 34(3): 911-925.
32. Yang X H, Luo Y Q, Yang H C, et al. Inversion and spatial distribution characteristics of soil salinity in oasis farmland in Manas River basin[J]. *Arid Zone Resources and Environment*, 2021, 35(2): 156-161.
33. Zhang L. Study on salinized land use change and utilization potential in oasis-desert area of Manas River Basin [D]. Xinjiang Agricultural University, 2013.
34. Gu G A. Formation of salinized soil and its prevention and control in Xinjiang [J]. *Xinjiang Geography*, 1984(4): 1-16.
35. Yang H C, Zhang F H, Wang D F et al. Trends of evapotranspiration from oases in the Mahe River Basin over the past 60 years and analysis of their influencing factors[J]. *Arid Zone Resources and Environment*, 2014,28(07):18-23.
36. Xin M L, Lv T B, He X L, et al. Spatial analysis of soil salinity in Manas River irrigation area based on ROC curve[J]. *Journal of Irrigation and Drainage*, 2016, 35(12): 45-50.
37. Tian F, Fensholt R, Verbesselt J, et al. Evaluating temporal consistency of long-term global NDVI datasets for trend analysis[J]. *Remote Sensing of Environment*, 2015, 163: 326-340.
38. Forkel M, Carvalhais N, Verbesselt J, et al. Trend Change Detection in NDVI Time Series: Effects of Inter-Annual Variability and Methodology[J]. *Remote Sensing*, 2013, 5(5): 2113-2144.
39. Vaudour E, Gomez C, Lagacherie P, et al. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2021, 96: 102277.
40. Shrestha R P. Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand[J]. *Land Degradation & Development*, 2006, 17(6): 677-689.
41. Alhammadi M S, Glenn E P. Detecting date palm trees health and vegetation greenness change on the eastern coast of the United Arab Emirates using SAVI[J]. *International Journal of Remote Sensing*, 2008, 29(6): 1745-1765.
42. Yao Y, Ding J L, Zhang F, et al. Regional soil salinization monitoring model based on hyperspectral index and electromagnetic induction[J]. *Spectroscopy and Spectral Analysis*, 2013, 33(6): 1658-1664.
43. Douaoui A E K, Nicolas H, Walter C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data[J]. *Geoderma*, 2006, 134(1-2): 217-230.
44. Abbas A, Khan S, Hussain N, et al. Characterizing soil salinity in irrigated agriculture using a remote sensing approach[J]. *Physics and Chemistry of the Earth, Parts A/B/C*, 2013, 55-57: 43-52.
45. Khan N M, Rastoskuev V V, Shalina E V, et al. Mapping Salt-affected Soils Using Remote Sensing Indicators - A Simple Approach With the Use of GIS IDRISI[J]. 2001.

46. Liu H J, Yang H X, Xu M Y et al. Soil classification based on multi-temporal remote sensing image features and maximum likelihood method during bare soil period[J]. *Journal of Agricultural Engineering*,2018,34(14):132-139+304.
47. Fu B L, Deng L C, Zhang L, et al. Remote sensing inversion of chlorophyll content in mangrove canopy with combined on-board hyperspectral imagery and stacked integrated learning regression algorithm[J]. *Journal of Remote Sensing*, 2022, 26(6): 1182-1205.
48. Zhang F, Li X, Zhou X, et al. Retrieval of soil salinity based on multi-source remote sensing data and differential transformation technology[J]. *International Journal of Remote Sensing*, 2023, 44(4): 1348-1368.
49. Zhang Z T, Tai X, Yang N, et al. Inversion of soil salinity by unmanned aerial vehicle multispectral remote sensing under different vegetation cover[J]. *Journal of Agricultural Machinery*, 2022, 53(8): 220-230.
50. Hu J, Lv YH. Progress in stochastic modeling of soil moisture dynamics[J]. *Progress in Geoscience*, 2015, 34(3): 389-400.
51. Chen H Y, Zhao G X, Chen J C, et al. Remote sensing inversion of saline soil salinity based on modified vegetation index in estuary area of Yellow River [J]. *Journal of Agricultural Engineering*, 2015, 31(5): 107-112+114+113.
52. Tang X L, Lv X. Impacts of climate change on available precipitation in the Manas River Basin over the past 50 years[J]. *Hubei Agricultural Science*, 2011, 50(22): 4582-4585.
53. Wei G, Li Y, Zhang Z, et al. Estimation of soil salt content by combining UAV-borne multispectral sensor and machine learning algorithms[J]. *PeerJ*, 2020, 8: e9087.
54. Zhang Z T, Wei G F, Yao Z H, et al. Research on soil salinity inversion modeling based on multi-spectral remote sensing by unmanned aircraft[J]. *Journal of Agricultural Machinery*, 2019, 50(12): 151-160.
55. Yang H, Hu Y, Zheng Z, et al. Estimation of Potato Chlorophyll Content from UAV Multispectral Images with Stacking Ensemble Algorithm[J]. *Agronomy*, 2022, 12(10): 2318.
56. Pham K, Won J. Enhancing the tree-boosting-based pedotransfer function for saturated hydraulic conductivity using data preprocessing and predictor importance using game theory[J]. *Geoderma*, 2022, 420: 115864.
57. Obsie E Y, Qu H, Drummond F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms[J]. *Computers and Electronics in Agriculture*, 2020, 178: 105778.
58. Li W D, Shi X Y, Song J H et al. Analysis of the dominant factors of soil physicochemical properties and salt ion composition in different geomorphic types of Manas River Basin[J]. *Journal of Shihezi University (Natural Science Edition)*,2022,40(01).
59. Zhang F H, Zhao Q, Pan X D, et al. Spatial differentiation of soil properties and rational development model of oasis in Mahe Basin, Xinjiang[J]. *Journal of Soil and Water Conservation*, 2005(6): 55-58.
60. Xia J, Wang S M, Zhu H W, et al. Spatial variability of soil salinity in the middle and lower reaches of the Manas River basin[J]. *Xinjiang Agricultural Science*, 2012, 49(3): 542-548.
61. Yan A, Jiang P A, Sheng J D, et al. Characterization of spatial variability of surface soil salinity in the Manas River Basin[J]. *Journal of Soil Science*, 2014, 51(2): 410-414.
62. Chen J H, Wang S M, Cao G D, et al. Physical properties of soils under different landforms and vegetation types in the Manas River Basin[J]. *Xinjiang Agricultural Science*, 2012, 49(2): 354-361.
63. Zhao Y C, HuDan T M E B, MaHeHuJiang A H M T et al. Characterization of intra- and inter-annual soil salinity changes in perennial drip-irrigated cotton fields in Northern Xinjiang[J]. *Research on arid region agriculture*,2015,33(05).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.