Estimating Land-Use Change Using Machine Learning: A Case Study on Five Central Coastal Provinces of Vietnam

Nguyen Hong Giang 1,2, Yu-Ren Wang 2,*; Tran Dinh Hieu 1,*; Nguyen Huu Ngu 3 and Thanh-Tuan Dang 4,5

1 Faculty of Architecture, Thu Dau Mot University, Thu Dau Mot 820000, Vietnam; giangnh@tdmu.edu.vn
2 Department of Civil Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807618, Taiwan
3 Faculty of Land Resources and Agricultural Environment, Hue University, Hue 49118, Vietnam; nhuungu@hueuni.edu.vn
4 Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 807618, Taiwan; i108143106@nkust.edu.tw
5 Department of Logistics and Supply Chain Management, Hong Bang International University, Ho Chi Minh 72320, Vietnam

* Correspondence: yrwang@nkust.edu.tw (Y.-R.W.); hieutd@tdmu.edu.vn (T.D.H.)

Abstract: Population growth is one factor relevant to land-use transformation and expansion in urban areas. This creates a regular mission for local governments in evaluating land resources and proposing plans based on various scenarios. This paper discussed the future trend of three kinds of land-use in the five central coast provinces. Afterwards, the paper deployed machine learning such as Multivariate Adaptive Regression Splines (MARS), Random Forest Regression (RFR), and Lasso Linear Regression (LLR) to analyze the trend of rural land use and industrial land-use to urban land-use in the Central Coast Region of Vietnam. The input variables of land-use from 2010 to 2020 were obtained by the five provinces of the Department of Natural Resources and Environment (DONRE). The results showed that these models provided pieces of information about the relationship between urban, rural, and industrial land-use change data. Furthermore, the MARS model proved to be accurate in the Quang Binh, Quang Tri, and Quang Nam provinces, whereas RFR demonstrated efficiency in the Thua Thien-Hue province and Da Nang city in the fields of land change prediction. Furthermore, the result enables to support land-use planners and decision-makers to propose strategies for urban development.

Keywords: Multivariate Adaptive Regression Spline (MARS); Random Forest Regression (RFR); Lasso Linear Regression (LLR); rural land-use; industrial land-use; urban land-use; decision-making

1. Introduction

Estimating land-use change provides valuable information about potential conversion that might occur over time in the earth’s surface landscapes. Furthermore, the prediction enables support for land-use planners and land resource managers to propose land-use change strategies, urban planning through modeling rural development, and selecting areas for setting industrial zones [1]; therefore, simulating models to observe and examine land-use change related to landscape dynamics over time is an interesting issue at both global and local scales [2].

Three major approaches are commonly applied in land-use change prediction, which are spatial pattern, statistical analysis, and artificial intelligence [3–13]. Models based on the simulation of the spatial pattern of land-use change processes, such as the Markov model, are deployed to know and interpret regional land changes and trends of regional land-use in effective ways [14]. Using machine learning based on quarterly or annual land-use statistics in localities to analyze land-use changes is also widely applied, such as Multivariate Adaptive Regression Splines (MARS) methodology, which is a kind of...
nonparametric (nonparametric covers techniques that do not rely on data belonging to any particular parametric family of probability distributions) and nonlinear technique used in statistical learning [15]. The Random Forest Regression (RFR) model also belongs to nonparametric learning, and the model is used in those areas [16,17]. The Lasso Linear Regression (LLR) model is the earliest form of least-squares prediction in classification, and its properties are similar to RFR and MARS [18–20]; hence, Yilmaz et al. (2018) [21] used MARS to estimate the suspended sediment load in Coruh River Basin, Turkey. The result showed that MARS outperformed the best model with R-squared is approximately 0.9, and they summarized that the MARS might be easily applied in modeling. Bui et al. (2019) [22] applied the MARS model to analyze and predict spatial patterns of forest fire danger for tropical forest fires in Lao Cai province, Vietnam. The result of the study pointed out that the model is the ability to solve the complexity of modeling forest fire danger. Nguyen et al. (2018) [23] deployed the RFR model and Landsat data for 10 classes consisting of multiple forest classes in Vietnam. The study result indicated that the overall accuracy is estimated at 0.90. Ha et al. (2020) [24] employed RFR and Landsat data with seven land-cover classes consisting of forest land to evaluate the land-cover classification in the northeast subtropical region of Vietnam. The study showed that the overall study accuracies were higher than 0.90. Denney-Frank et al. (2019) [25] used LLR with cross-validation forecasting streamflow impacts of forest restoration and conservation based on simulation of the hydrology of 29 located models worldwide. The result demonstrated that the model for water yield change after the development of agriculture with R-squared is around 0.69 when using LLR model.

Although not many types of research have been announced to compare methods correlated to accuracy levels land-use changes in the five central coastal provinces of Vietnam; therefore, deploying these three proposal models to estimate urban land-use change for this region is feasible and may provide a high forecasting accuracy.

The Quang Binh, Quang Tri, Thua Thien-Hue, Da Nang, and Quang Nam provinces belong to the Central Coast Region of Vietnam that has made positive changes in the process of urbanization in recent years. As a result, there are a lot of larger urban areas in the region that are formed in the eastern coastal area; however, the process of urbanization in the region has revealed significant shortcomings, limitations, and challenges.

This study aims to present the estimation of the urban land-use change using LLR, RFR, and MARS models. The input vectors used in the models are based on the land-use change as rural land-use, industrial land-use, and urban land-use 44 quarters in five central coastal provinces in Vietnam between 2010 and 2020. With the support of the collected data, the paper discusses the role of three types of land-use in the region’s urbanization process. After that, it highlights a comparison between three models based on statistical accuracy indicators’ results. Furthermore, the collection of results of these three models may show the working efficiency of the models for land-use prediction, and it may develop future scenarios that can support land-use planning and decision-making.

The structure of the paper is organized as follows. Section 1 gives the paper introduction. Section 2 introduces the materials and methodology of MARS, RFR, and LLR models. Sections 3 and 4 describe the results and discussions. Finally, Section 5 presents the conclusions.

2. Materials and Methods

The process of the following experimental stages in this study is described in Figure 1. Firstly, the database of urban, rural, and industrial land use is preprocessed and tested by statistical methods in the input layer. Secondly, the database is divided into 70% for the training phase and 30% for the testing phase, and the MARS, RFR, and LLR models are used to learn the training samples and obtain the optimal network parameters during the process. Finally, the three models’ implementations are showed the base functions and are compared using metrics from the accuracy measurement indicators as RMSE, MAE, MSE,


R, R² in the possible result stage, at the same time, looking for the most suitable prediction model for the study.

![Diagram of the research steps used in this study.](image)

**Figure 1.** Diagram of the research steps used in this study.

### 2.1. Study Area

The Middle Central Coast region comprises five provinces and cities that serve as a link between the country’s three primary socioeconomic hubs: the Red River Delta, the South Central Coast, and the Central Highlands (see Figure 2). The topography of this area is divided into two parts: the flat to gently rolling plains in the east and the most rugged forest-covered mountains in the western two-thirds. The delta area has a coastal-mountainous nature, and it is separated by mountain branches close to the sea as Hoanh Son mountain range—Ngang pass, Bach Ma mountain range—Hai Van pass [26,27]. Furthermore, this region plays a critical role in Vietnam’s maritime economic development strategy in terms of tourism development, research, and technology, the Vietnamese seaport system, and vital logistics. When it comes to administrative matters, the North Central region is divided into six provinces with a total size of 2,948,430 hectares (9.8% of the country’s total area) and a population of 5,366,500 people (5.70 percent of the total population); the rates of average population growth of the region is 1.1% [28]. These points show that the urbanization rate in this area is not high, and the population growth rate is only low compared to Hanoi and Ho Chi Minh City regions. Furthermore, there are two first-grade cities in the region (Da Nang and Hue), two second-grade cities (Tam Ky and Dong Hoi), two third-grade cities (Dong Ha and Hoi An), and eight fourth-grade cities in total, and 35 fifth-grade cities (see Appendix A Table A1) [29–33].

On the other hand, the region’s average urbanization rate is 47.26 percent, with Da Nang having the highest rate at 87 percent and Quang Binh having the lowest at roughly 30 percent. Following that, agriculture, forestry, fisheries, and industry—construction, services, and product tax minus product subsidies accounted for 15%, 28%, 48%, and 9% of the region’s GRDP in 2021, respectively. Per capita income based on GRDP is $2,643 USD/person, with Da Nang having the most at $3,822 USD/person and Quang Tri province having the lowest at $2,087 USD/person. As this field expands, vocational training is designed to help people change occupations. Feeding a portion of the rural population and those who no longer have fertile land is a critical challenge for this region. By 2021, those with vocational training will make up approximately 66 percent of the working-age population. Their earnings will rise because they are well-educated, lowering the poverty rate to around 3.5 percent in 2021 [29–33]; therefore, vocational training demonstrates that urbanization stimulates and expands children’s opportunities. People are more dynamic and innovative in their search for, and selection of, methods and forms of production, and organizations rise to become wealthy legally. The main trend and best aspect of urbanization is economic development, which improves employees’ living standards. The
rapid development of non-manufacturing industries is also aided by urbanization. Large cities also provide more work options, greater pay, better social services, and increased labor productivity. It is a driving force for economic transformation in both urban and rural areas, contributing to further economic development. At the same time, the metropolitan region serves as a big and diverse consumer of goods, a location to employ a skilled workforce and a hub for sophisticated technology and infrastructure facilities that draw significant domestic and foreign investment.

Urbanization in these provinces is established mainly in two ways, as follows. Firstly, the land-cover transforms rural areas into urban areas, then the villages and communes surrounding urban centers are gradually merged into urban areas. Alternatively, some rural areas have developed enough infrastructure, and the total population of the rural region meets the criterion of an urban population (higher than 50,000 people), that the place will be recognized as an urban area. Lastly, the developing industrial, commercial, and tourism zones promote neighboring suburb areas to develop as urban areas.

The total land-use of rural, urban, and industry areas within these provinces estimate about 71,067 ha in 2020 (see Table 1), in which Quang Binh, Quang Tri, Thua Thien-Hue, Da Nang, and Quang Nam occupy 9973 ha, 6341 ha, 14,510 ha, 11,834 ha, and 28,409 ha, respectively. In addition, Figure 3 indicates the square of land-use change in this study scope from 2010 to 2020. Figure 3a shows that the most change in urban land-use occurs in the Thua Thien Hue province, followed by Da Nang city, and the lowest variation is in the Quang Binh and Quang Tri provinces. Figure 3b indicates that the variation in rural land-use use in the Quang Nam province is the most, followed by the change in Thua Thien Hue and Quang Binh provinces, with the lowest variation occurring in the Da Nang and Quang Tri provinces. Finally, Figure 3c points out that the change in land-use of industrial zones in Quang Nam province is the highest, followed by Da Nang, then Thua Thien Hue, with the lowest being in the Quang Binh and Quang Tri provinces.

Figure 2. Location of provinces for study.
Table 1. The area of land-use categories by province in the study area in 2020 (Unit: ha).

| Province        | Rural Land-Use (ha) | Industrial Land-Use (ha) | Urban Land-Use (ha) | Sub-Total |
|-----------------|---------------------|--------------------------|---------------------|-----------|
| Quang Binh      | 5632                | 3103                     | 1238                | 9973      |
| Quang Tri       | 3067                | 1740                     | 1534                | 6341      |
| Thua Thien-Hue  | 6420                | 4596                     | 3494                | 14,510    |
| Da Nang         | 2464                | 4694                     | 4676                | 11,834    |
| Quang Nam       | 17,024              | 6751                     | 4634                | 28,409    |
| Total           | 34,607              | 20,884                   | 15,576              | 71,067    |

Figure 3. The area of (a) urban land-use (unit: ha), (b) rural land-use (unit: ha), and (c) industrial land-use (unit: ha) of five provinces from 2010 to 2020.
2.2. Database

This paper contains a database containing 44 quarters of land-use change from 2010 to 2020 (4 quarters for each year) for five provinces obtained from the Department of Natural Resources and Environment (DONRE). Three input variables include rural land-use, industrial land-use, and urban land-use, which were collected from five provinces. Furthermore, the characteristic statistical results for urban land-use in Table 2 demonstrate that the mean ranges from 878 ha in Quang Binh to 4317 ha in Da Nang, the standard deviation (St Dev) ranges from 84 ha in Quang Tri to 959 ha in Da Nang, and the minimum (Min) and maximum (Max) range from 608 ha and 1238 ha in Quang Binh to 4093 ha and 4634 ha in Quang Nam. The ranges of skewness (Skew) and kurtosis (Kurt) parameters of the five provinces fluctuate from 0.17 and −1.72 to 1.03 and 1.65. These Skew and Kurt indicators approach low values that are highly appropriate for modeling [34]. The input data patterns of five provinces were randomly selected in two parts. About 70% of the dataset was selected for the training sample, whereas 30% was used for the testing sample.

| Province   | St Dev (ha) | Mean (ha) | Min (ha) | Max (ha) | Skewness | Kurtosis |
|------------|-------------|-----------|----------|----------|----------|----------|
| Quang Binh | 214         | 878       | 608      | 1238     | 0.17     | −1.17    |
| Quang Tri  | 84          | 1369      | 1262     | 1534     | 0.64     | −0.86    |
| Thua Thien-Hue | 959     | 4076      | 3272     | 5434     | 0.58     | −1.72    |
| Da Nang    | 403         | 4317      | 3514     | 4676     | 1.03     | −0.67    |
| Quang Nam  | 183         | 4219      | 4093     | 4634     | 0.87     | 1.65     |

2.3. Descriptions of Models

2.3.1. Multivariate Adaptive Regression Splines (MARS)

MARS is a nonparametric regression model, and it was introduced by Friedman [35]. MARS seems like a method for a fitted relationship between prediction and dependent variables. MARS is fast and based on a divide-and-conquer strategy, which divides the training dataset into distinct regions, each with its regression line [36–38]. The MARS algorithm feature is the procedure of the backward and forwards stepwise and may explain and control the complex nonlinear mapping between the inputs and output variables. This function predicts the new output \( y \) and the input variable \( x \) that uses either of the two base functions [39] and deploys a value or knot of variables that demonstrates the point of inflection along with the range of the inputs [40]. The general form of MARS forecasting is as below:

\[
y = f(x) = \beta_0 + \sum_{j=1}^{P} \alpha_j \beta_j
\]  

where \( y \) is the dependent variable predicted by the function \( f(x) \); \( \beta_0 \) is the constant value; \( P \) is the number of terms, each of them formed by a coefficient \( \alpha_j, j \in \{1, \ldots, P\} \); \( x \) is predictor variable; \( \beta_j \) is an individual base function. The base functions of \( \text{Max}(0, x - H) \) and \( \text{Max}(0, H - x) \) are univariate and do not have to each be present if their \( \beta \) coefficients are 0; the \( H \) values are called “hinges” or “knots”; \( x \) is an independent variable.

The function of backward stepwise relates to removing basis functions one at a time until the criterion of “lack of fit” is a minimum. In the deleting stage of backward stepwise, the last crucial important base functions are demolished one at a time. The lack of an applied fitting measurement is leaned on in the Generalized Cross-Validation (GCV) [41,42]:

\[
\text{GCV} = A * \sum_{i=1}^{P} \frac{(y_i - \hat{f}(x))/N}{N}
\]
where \( N \) is a number of data; \( A = \left[ 1 - \frac{C(M)}{N} \right]^{-2} \) and \( C(M) = (M+1) + dM \) are the complexity function \([35]\); \( d \) is a penalty for each basis function included in the model; \( M \) is the number of base functions in Equation (1). The criterion of GCV is examined for the average residual error multiplied by a penalty to modify the variability associated with more indicator prediction in the model \([39,43]\).

2.3.2. Lasso Linear Regression (LLR)

The lasso linear regression method is widely used in domains with massive datasets, and it is also necessary to use when algorithms are efficient and quick \([44]\); however, the lasso is not vigorous in terms of determining the high correlation between predictors; it will randomly choose one and ignore the others, and split when all predictors are identical \([44]\). Moreover, the lasso penalty looks at many coefficients that are close to zero and only a small subset that is larger (and non-zero). The lasso estimator \([45,46]\) uses the \( l_1 \) penalized least-squares criterion to get a sparse solution to the problem of optimization as below:

\[
\hat{\beta}_{\text{Lasso}}(\beta) = \arg\min_{\beta} \|y - X\beta\|_2^2 + \gamma \|\beta\|_1
\]

(3)

where \( \|\beta\|_1 = \sum_{j=1}^{p} |\beta_j| \) is the \( l_1 \)-norm penalty on \( \beta \), which is the cause of the sparse solution, and \( \gamma \geq 0 \) is a tuning parameter.

The \( l_1 \) penalty allows the lasso to simultaneously fit the smallest squares to and shrink some components of \( \hat{\beta}_{\text{Lasso}} \) to zero for a suitably chosen \( \gamma \) \([44]\). The cyclic coordinate reduction algorithm \([44]\) efficiently computes the entire path of the Lasso solution paths for \( \gamma \) for the Lasso estimator and is faster than the Generalized Least Angle Regression (LARS) well-known algorithm. These properties make Lasso an attractive and popular method of variable selection.

2.3.3. Random Forest Regression (RFR)

Random forest is a regression technique that associates the performance of multiple decision tree algorithms to classify or forecast the value of a variable \([47–49]\). When an \( x \) input vector is received by random forest, in conjunction with the different evidential features values analyzed for a given training area, a number \( K \) of regression trees, on averages, the results are built by random forest. After \( K \), such trees \( \{T(x)\}_{K=1}^{K} \) are grown, and the random forest regression predictor is as follows:

\[
\hat{f}_{\text{rf}}(x) = \frac{1}{K} \sum_{k=1}^{K} T(x)
\]

(4)

To avoid the correlation of different trees, the random forest raises the tree’s diversity by improving from different subsets of training data generated through a procedure called bagging \([50]\). Bagging is a technique used to generate training data by randomly resampling the original dataset with a replacement. As a result, some data may be used multiple times during training, whereas others may never be used. Thus, greater stability is achieved, as it makes it more robust in the face of small variations in the input data, and at the same time, it increases the accuracy of the prediction \([47,51]\).

2.3.4. Performance Metrics

Predicting results is based on calculating and comparing the actual values to the forecasted values. These metrics of the accuracy measurement parameters include the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE),
Correlation Coefficient (R), and Correlation of Determination ($R^2$). Furthermore, the error metrics are defined as follows [52–54]:

\[
\text{MSE} = \frac{\sum_{t=1}^{n} (x_t - x'_t)^2}{n} \quad (5)
\]

\[
\text{MAE} = \frac{\sum_{t=1}^{n} |x_t - x'_t|}{n} \quad (6)
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} (x_t - x'_t)^2}{n}} \quad (7)
\]

\[
R^2 = 1 - \frac{\sum_{t=1}^{n} (x_t - x'_t)^2}{\frac{1}{n} \sum_{t=1}^{n} x_t^2} \quad (8)
\]

\[
R = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(x'_t - \bar{x})}{\sqrt{\sum_{t=1}^{n} (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^{n} (x'_t - \bar{x})^2}} \quad (9)
\]

where $x_t$, $x'_t$ are the observed and estimated values in the period time $t$, and $n$ is the number of the observed values in the testing data. $\bar{x}$, $\bar{x}'$ are mean of the observed and estimated value. The $R^2$ and $R$ (correlation coefficient) should be approaching 1 to indicate strong model performance, and the MSE, MAE, and RMSE should be as close to zero as possible.

3. Results Analysis

Regarding this platform, the output function of MARS and LLR of land-use in five provinces are presented as below:

**Quang Binh (QB) land-use output function:**

MARS$_{\text{Quang Binh}} = 990 + 1.27F_{1QB} - 0.55F_{2QB} - 0.18F_{3QB} + 0.65F_{4QB} - 0.52F_{5QB} + 0.27F_{6QB}$, where $F_{1QB} = \max(0, \text{Rural-5424})$, $F_{2QB} = \max(0, 5424-\text{Rural})$, $F_{3QB} = \max(0, \text{Industry-2251})$, $F_{4QB} = \max(0, 2251-\text{Industry})$, $F_{5QB} = \max(0, \text{Industry-2995})$, and $F_{6QB} = \max(0, \text{Industry-2366})$.

\[
\text{LLR}_{{\text{Quang Binh}}} = -2563 + 0.7X_{1QB} - 0.1X_{2QB}.
\]

$F_{iQB}$ ($i = 1, 2, \ldots, 6$) is the base function. $F_{iQB}$ may be explained as the maximum value of 0 and Rural-5424. The minus sign ahead of the maximum value is equivalent to a minimum value. In addition, the MARS$_{\text{Quang Binh}}$ analysis indicates that the most important is in bellowing order rural land-use and industrial land-use.

**Quang Tri (QT) land-use output function:**

MARS$_{\text{Quang Tri}} = 1435 + 0.94F_{1QT} + 0.45F_{2QT} + 1.57F_{3QT} - 2.45F_{4QT} - 0.58F_{5QT}$, where $F_{1QT} = h(\text{Industry-1285})$, $F_{2QT} = \max(0, 1285-\text{Industry})$, $F_{3QT} = \max(0, 3047-\text{Industry})$, $F_{4QT} = \max(0, 3047-\text{Rural})$, and $F_{5QT} = \max(0, \text{Industry-1185})$.

\[
\text{LLR}_{{\text{Quang Tri}}} = -308 + 0.42X_{1QT} + 0.3X_{2QT}.
\]

The MARS$_{\text{Quang Tri}}$ analysis indicates that the most important is in bellowing order industrial land-use and rural land-use.

**Thua Thien-Hue (TTH) land-use output function:**

MARS$_{\text{Thua Thien-Hue}} = 1718 + 14.97F_{1TTH} + 12.88F_{2TTH} - 2.91F_{3TTH} - 12.79F_{4TTH}$, where $F_{1TTH} = \max(0, 6277-\text{Rural})$, $F_{2TTH} = \max(0, 3428-\text{Industry})$, $F_{3TTH} = \max(0, 3428-\text{Industry})$, $F_{4TTH} = \max(0, \text{Industry-3558})$.

\[
\text{LLR}_{{\text{Thua Thien-Hue}}} = 105,973 - 18.43X_{1TTH} + 3.40X_{2TTH}.
\]

The MARS$_{\text{Thua Thien-Hue}}$ analysis indicates that the most important is in bellowing order rural land-use and industrial land-use.
Da Nang (DN) land-use output function:
MARS_{DaNang} = -3086 + 1.18F_{1DN} - 1.11F_{2DN} + 1.72F_{3DN}, \text{where } F_{1DN} = \max(0, \text{Industry}-4000), F_{2DN} = \max(0, \text{Industry}-4408), F_{3DN} = \max(0, \text{Industry}).

LLR_{DaNang} = -1280 + 0.34X_{1DN} + 1.08X_{2DN}.

The MARS_{DaNang} analysis indicates that the most important is in bellowing order industrial land-use and rural land-use.

Quang Nam (QN) land-use output function:
MARS_{Quang Nam} = 4215 - 0.5F_{1QN} - 0.11F_{2QN} - 0.17F_{3QN} + 0.06F_{4QN}, \text{where } F_{1QN} = h(\text{Industry}-5922), F_{2QN} = \max(0, 5922-\text{Industry}), F_{3QN} = \max(0, 16,532-\text{Rural}), \text{and } F_{4QN} = \max(0, 16,532-\text{Rural}).

LLR_{Quang Nam} = -1651 + 0.29X_{1QN} + 0.19X_{2QN}.

The MARS_{Quang Nam} analysis indicates that the most important is in bellowing order industrial land-use and rural land-use.

Furthermore, the output function for RFR does not occur.

The data in Figures 4a–c, 5a–c, 6a–c, 7a–c and 8a–c present the relationship between the three types of land-use, and the area of industrial and rural land-use increases as the square of urban land-use increases. LLR makes a forecasting form that resembles a flat surface of paper. Additionally, the RFR and MARS charts are the same as the image of papers with some folds, and the folds enable a better fit to the data. In addition, Figures 4a–c, 5a–c and 8a–c demonstrate that the area of land usage in the Quang Nam province increased significantly from 2010 to 2020. Moreover, the MARS predicted algorithm shows that the red dots are evenly distributed on the surface, and the LLR and RFR forecasting algorithms demonstrate that the red dots are relatively far from the surfaces. The data in Figure 6a–c imply that the area used for the three categories of land experienced an upward tendency between 2010 and 2013; however, the urban land use area decreased dramatically and tended to be saturated, whereas the rural land use area and industrial zones grew steadily per annum from 2014 to 2020. The output in Figure 7a–c illustrates that the urban and industrial land use increased significantly from 2010 to 2020, whereas rural land use grew slowly and even reached its saturation in 2019. Additionally, the RFR forecasting model of Figures 6a and 7a shows that these red dots are moderately distributed closer to the surface compared with the two algorithms in Figures 6b,c and 7b,c; however, it is difficult to investigate the difference between the three models. Hence, this study performed an accurate metric to explore the potential models for each province. From comparing the three models, Table 3 shows that the MARS model supplies a better fit than the other models for Quang Binh, where $R^2$ value = 0.91 was the largest along with MSE and MAE, whereas RMSE obtained the lowest value out of the established models. According to the implementation of the other models, the hierarchical carrying out considers the order of MARS > RFR > LLR. The simultaneous determination of urban land-use change prediction of Quang Nam and Quang Tri also proves the forecasting skills of these models in urban land-use change prediction in these provinces. Using the experiment results from the three provinces, the MARS model indicated to supply the best forecasting accuracy, the hierarchical order of the models, and other models for three provinces are MARS > RFR > LLR in Quang Tri and Quang Nam. Moreover, the prediction result of urban land-use change prediction in Thua Thien-Hue and Da Nang presented in Table 3 denotes that the order of hierarchical models with performance accuracy is RFR > MARS > LLR models. Regarding the GVC parameter for MARS, it generates an equilibrium between flexibility and generalization capability of the MARS model function [55]. The data in Table 3 also indicates that the order of hierarchical models with the accuracy of the GVC value of the MARS model in five provinces are GVC_{Thua Thien-Hue} > GVC_{Da Nang} > GVC_{Quang Tri} > GVC_{Quang Binh} > GVC_{Quang Nam}. Furthermore, the scatter charts in Figure 9a,b,e proved that the line of MARS models (red plus lines) of urban land used change estimations fit better than the line
of RFR models (green triangle lines) and the line of LLR models (purple multiply lines) for the Quang Binh, Quang Tri, and Quang Nam provinces. In contrast, Figure 9c,d showed that the line of RFR models of urban land used change prediction in Thua Thien-Hue which showed that Da Nang has the best fit, followed by the line of MARS models and line of LLR models, respectively. These points may explain that the distribution of land-use change data with random selection for training and testing data is suitable for the MARS model of Quang Binh, Quang Tri, and Quang Nam, and is comfortable for the RFR model in Thua Thien-Hue and Da Nang.

![Image of land-use prediction with (a) RFR model, (b) MARS model, and (c) LLR model in Quang Binh (unit: ha).](image)

Table 3. Accuracy parameters for land-use prediction.

| Quang Binh | Quang Tri | Thua Thien-Hue | Da Nang | Quang Nam | Average |
|------------|-----------|----------------|---------|-----------|---------|
| MSE        | 143       | 10             | 5       | 107       | 10      | 107     | 106     | 105       | 104       | 103       | 100       | 101       | 100       | 100       |
| MAE        | 27        | 4              | 5       | 33        | 6        | 6       | 442     | 70        | 162       | 78       | 79       | 75        | 79        | 78        | 78        | 78        |
| RMSE       | 38        | 11             | 6       | 39        | 9        | 9       | 535     | 150       | 131       | 98       | 98       | 84        | 98        | 98        | 98        | 98        |
| R          | 0.91      | 0.91           | 0.91    | 0.82      | 0.91     | 0.91    | 0.67    | 0.92      | 0.91      | 0.89     | 0.87     | 0.89      | 0.91      | 0.91      | 0.91      | 0.91      |
| R²         | 0.92      | 0.94           | 0.94    | 0.66      | 0.93     | 0.93    | 0.56    | 0.92      | 0.92      | 0.9     | 0.91     | 0.91      | 0.91      | 0.91      | 0.91      | 0.91      |
| GCV        | 88        | 193            | 66,746  | 7522      | 77       | 77      | 77      | 77        | 77        | 77       | 77       | 77        | 77        | 77        | 77        | 77        |
Figure 5. Land-use prediction with (a) RFR model, (b) MARS model, and (c) LLR model in Quang Tri (unit: ha).
Figure 6. Land-use prediction with (a) RFR model, (b) MARS model, and (c) LLR model in Thua Thien-Hue (unit: ha).

Figure 7. Cont.
Furthermore, the predicting performance of the model models is also visualized and examined by the Taylor diagram. The diagram summarizes the St Dev and correlation coefficient (CC) that is comprised concomitantly in assessing the respective model [56,57]. The St Dev, and CC between the observed and predicted datasets for all the land-use

Figure 7. Land-use prediction with (a) RFR model, (b) MARS model, and (c) LLR model in Da Nang (unit: ha).

Figure 8. Land-use prediction with (a) RFR model, (b) MARS model, and (c) LLR model in Quang Nam (unit: ha).
models of the provinces are described in the Taylor diagram. Figure 10a–e may be observed for LLR, RFR, and MARS in Quang Binh (CC_{LLR} = 0.91, CC_{RFR} = 0.91, CC_{MARS} = 0.91), in Quang Tri (CC_{LLR} = 0.82, CC_{RFR} = 0.91, CC_{MARS} = 0.91), in Thua Thien-Hue (CC_{LLR} = 0.67, CC_{RFR} = 0.92, CC_{MARS} = 0.91), in Da Nang (CC_{LLR} = 0.89, CC_{RFR} = 0.87, CC_{MARS} = 0.89), and in Quang Nam (CC_{LLR} = 0.86, CC_{RFR} = 0.91, CC_{MARS} = 0.92). The Taylor diagram demonstrates that these models were optimal accuracies of almost all models’ outcomes and were significantly closer to 1. Moreover, the LLR model of land-use in Thua Thien-Hue with CC_{LLR} = 0.67 indicates that the level of accuracy achieved is only above medium.

**Figure 9.** The best performance models for urban land use change prediction; (a) QuangBinh province, (b) QuangTri province, (c) ThuaThienHue province, (d) Da Nang City, (e) QuangNam province.
4. Discussion

Urban areas in the five central coastal provinces are organized evenly along the coast, primarily on urban beam space. Rural areas still account for a much larger proportion of land than urban areas, and approximately 15% to 20% of the land use belongs to the urban administrative boundary, and the population accounts for over 60%. The urbanization process creates a sharp change in land-use in peri-urban and rural areas. The conversion of a large part of agricultural land to land for the construction of industrial, service, and urban residential areas. The process of expanding urban space along with the appearance of housing projects, real estate, concentrated industrial parks, large-scale commercial service works in the peri-urban communes has caused a sharp decline in production land funds, natural land, and spatial change of the rural ecological landscape, which causes the mechanical population of peri-urban communes to increase, despite insufficient infrastructure, leading to rapid overcrowding. Moreover, it also has an impact upon technical infrastructure, social infrastructure, especially traffic, education, water supply, and drainage, environmental sanitation; however, urbanization has lagged behind the growth of industrial zones, and industrial development lacks a vision for future urbanization. In industrial zones, a large number of workers leaving the agricultural production area tend to move to the industrial production area, forming areas with high population density, centralization, creating demand for services such as food, accommodation, living, studying, and commuting purposes.
which are the premise for the initial formation of an industrial residential area, an industrial town, and in the future, it will become an industrial city. Hence, the following are the primary types of the relationship between the growth of industrial parks and the process of urbanization. Firstly, many industrial parks are located in rural areas but not in urban areas. Secondly, several industrial parks were formerly located in rural areas, but now it is still within urban areas’ boundaries. Following that, there are many industrial parks in rural areas, which are now within the proposed expansion boundaries of the neighboring urban master plan. Consequently, the above situation shows that assessing the influence of rural land-use and industrial zone land-use is significantly vital for urban land-use and urbanization in this area.

This study implemented three machine learning models for land-use change to assess the speed of urbanization taking place in the Central Coast Region in Vietnam. Three models gave high accurate results for predicting urban land-use fluctuations, in which the MARS and RFR models showed more accuracy for Quang Binh, Quang Tri, Quang Nam, and Thua Thien-Hue, Da Nang, respectively, compared with the LLR model. In addition, the estimated values of types of land-use changes made by the LLR model also provided acceptable results. The data of this study was based on the statistics of land-use types that have been measured quarterly; therefore, these estimation values supplied the total types of land-use change that have been urbanized based on the process of forming urban areas in industrial and rural areas.

To evaluate the predicted accuracy of these study models, spatial models are needed to estimate accuracy parameter values. Comparing RMSE and MAE using the MARS model of this study result with the study result of Yilmaz et al. (2018) [21] about suspended sediment load, their RMSE = 3592, and MAE = 3483 are found to be greater than the values in this study with RMSE\textsubscript{Average} = 47.9, MAE\textsubscript{Average} = 50.8. Jamali (2019) [58] deployed RFR to predict land-use/land-cover mapping using Landsat 8 OLI in the northern region of Iran. The RMSE and MAE for the model are 5 and 5, respectively; these points are lower than this result study. Finally, the result of the study Adab et al. (2020) [59] concerns Estimate Surface Soil Moisture in the semi-arid region of west Khorasan-Razavi province of Iran, and it shows that RMSE and MAE of LRR model (at 7 March 2017) are 6.67 and 5.55, respectively. These points also indicate that the prediction model is lower than this study model. Duong et al. (2018) [60] deployed the kernel density estimation and remotely sensed data from multiple sensors to generate the land cover maps over Central Vietnam during the period of 2007 to 2017. The result indicated that the overall accuracies of the maps for 2007 and 2017 are 90.5% (kappa coefficient of 90%) and 90.6% (kappa coefficient of 90%), respectively, in which the urban prediction was approximately 91%. This point also proved that using machine learning to show the results of this study consider equivalent to the remote sensing method for estimating the land use/land cover for the Center of Vietnam; however, the study deploys machine learning models and statistical algorithms that majorly focus on land use transition/change. Due to sufficient published literature relating to other aspects of land use planning such as zoning, land allocation, and land restrictions, land-use mapping was not mentioned in this study; therefore, combining multiple methods for land-use information would be useful in future research.

Although classification accuracies for land-use were not particularly large, estimating urban, rural, and industrial land-use change is still useful for five central coastal provinces of Vietnam. This study result will assist the provinces’ authorities and other stakeholders in decision-making and planning regarding three kinds of land-use. The usual practice is for the Ministry of Natural Resources and Environment (MONRE) to carry out urban inventory and set up urban land-use change maps every five years. Then, the DONRE provinces obtain the predictive data and update them manually. In addition, many jobs are created as a result of the development of the service, commerce, and manufacturing industries in cities. At the same time, a lot of individuals lose arable land due to urbanization to make way for industrial parks, handicrafts, or concentrated craft villages. Moreover, urbanization will affect policymakers regarding labor reorganization, changing production methods,
and enhancing human resource training solutions to adapt to new employment standards. Moreover, new industries and services drive economic growth. Furthermore, sustainable urbanization development is a concern, and several criteria need to be proposed as below. Firstly, harmonious development of the economy, society, environmental protection, and ecological balance is required. Secondly, the municipality must ensure that the amount of space available for activities, the infrastructure engineering system, and social infrastructure are all up to par with high-quality standards. Thirdly, cities must have a well-organized population distribution system to close the gap between urban, rural, and industrial zones. Fourthly, urban development must balance the ecology in the inner city and suburbs. Finally, the authorities have to enforce appropriate policies related to population, land use, technical infrastructure development, environmental protection, and preservation of natural and social ecosystems.

Therefore, sustainable urban development for urban provinces can be suggested as follows:

1. Da Nang city, the most developed urban area in the region, has industrial parks equivalent to the urban land-use area. As a result, it is critical to relocate industrial zones in the ancient city, rationalize land use functions, employ high-tech equipment, and create a green environment. More importantly, to accommodate the influx of migrants from all over the country into the city’s working streets, local authorities must plan to build land funding and infrastructure in industrial zones or rural areas near industrial zones, lowering stress in Da Nang’s central city.

2. The Quang Nam and Thua Thien Hue provinces, two provinces with many tangible cultural heritage sites such as Hue City, Hoi An Ancient Town, and My Son Holyland, need to build satellite urban areas to relieve the pressure on infrastructure and population for urban heritage areas. In addition, because the land fund for rural use is enormous, a strategy for converting agricultural land to industrial and commercial zones in rural areas is required to support rural growth and urban areas while also creating jobs for rural residents.

3. Quang Binh and Quang Tri are two provinces with slower urban and industrial zone development than the Da Nang, Quang Nam, and Thua Thien-Hue provinces; however, plenty of rural land use and industrial land use funds are being used in these two provinces. As a result, these two provinces will need to construct satellite cities based on highly populated areas near industrial parks. In addition, it is necessary to form sub-regional centers in the district in the direction of commodity production with high technology.

5. Conclusions

Urbanization is an inevitable process for the economic and social development of the five central coastal provinces; therefore, this study has shown the role of rural land-use and industrial zone land-use in informing and expanding urban areas. However, this development has not been synchronized and has not yet ensured the infrastructure of an urban area; hence, this study used the MARS, RFR, and LLR models to estimate urban land-use change based on rural and industrial land-use from the five provinces. In addition, the projected and observed values were compared using five widely used statistical parameters (i.e., RMSE, MAE, MSE, R, and $R^2$). The results of the study of the models also show that the MARS model improves the accuracy of performance more than RFR, LLR in the Quang Binh, Quang Tri, Quang Nam provinces, and the RFR model gives a more accurate forecasting implementation than the MARS, LLR models in the Da Nang and Thua Thien-Hue provinces. The accuracy of the models may depend on the distribution of land-use change data with random selection for training and testing data. The prediction of land-use change may support the authorities’ land-use planning and decision-making. Furthermore, the research also suggested sustainable urban development for each specific province, and the region in general. The future of the current work consists of using a hybrid of MARS and LLR in modeling land-use change and other studies to enhance the
model estimating capability, or these methods may combine with the spatial pattern models to estimate land-use change.

**Author Contributions:** Conceptualization, N.H.G.; data curation, N.H.N.; formal analysis, Y.-R.W.; funding acquisition, N.H.G.; investigation, Y.-R.W.; methodology, N.H.G.; project administration, Y.-R.W.; software, N.H.G.; validation, Y.-R.W.; writing—original draft, N.H.G.; writing—review and editing, T.D.H. and T.-T.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors appreciate the support from the National Kaohsiung University of Science and Technology, Taiwan; and Thu Dau Mot University, Vietnam.

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

**Appendix A**

**Table A1.** Urban classification of Vietnam.

| Criteria/Indicators       | Type I                              | Type II                                 | Type III                             | Type IV                              | Type V                              |
|---------------------------|-------------------------------------|-----------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|
| Population                | (a) >1 million: Central government-run city | (c) 300,000 to 1 million: If class 2 is central government-run city, population should be more than 800,000 | (d) 100,000 to 350,000               | (e) 500,000 to 350,000               | (f) >4000                           |
|                           | (b) 500,000: Provincial city          |                                         |                                      |                                      |                                     |
| Nonagricultural labor     | 85%                                 | 80%                                     | 70%                                  | 70%                                  | >65%                                |
| Population density        | (a) 12,000/km²                       | 8000 /km² or 10,000 /km² if the city is directly under central government control | 600 km²                             | 4000 km²                             | 2000 km²                            |
|                           | (b) 10,000/km²                       |                                         |                                      |                                      |                                     |

**References**

1. Pijanowski, B.C.; Brown, D.; Shellito, B.A.; Manik, G.A. Using neural networks and GIS to forecast land use changes: A Land Transformation Model. *Comput. Environ. Urban Syst.* 2002, 26, 553–575. [CrossRef]

2. Sang, L.; Zhang, C.; Yang, J.; Zhu, D.; Yun, W. Simulation of land use spatial pattern of towns and villages based on CA–Markov model. *Math. Comput. Model.* 2011, 54, 938–943. [CrossRef]

3. Wang, S.W.; Gebru, B.M.; Lambin, M.; Kayastha, R.B.; Lee, W.-K. Land Use and Land Cover Change Detection and Prediction in the Kathmandu District of Nepal Using Remote Sensing and GIS. *Sustainability* 2020, 12, 3925. [CrossRef]

4. Leping, C.; Yujun, S.; Saeed, S.; Saeed, S. Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques—A case study of a hilly area, Jiangle, China. *PLoS ONE* 2018, 13, e0200493. [CrossRef]

5. Saputra, M.H.; Lee, H.S. Prediction of Land Use and Land Cover Changes for North Sumatra, Indonesia, Using an Artificial-Neural-Network-Based Cellular Automaton. *Sustainability* 2019, 11, 3024. [CrossRef]

6. Veldkamp, A.; Lambin, E. F. Predicting land-use change. *Agric. Ecosyst. Environ.* 2001, 85, 1–6. [CrossRef]

7. Viney, N.R.; Bormann, H.; Breuer, L.; Bronstert, A.; Croke, B.; Frede, H.; Graf, T.; Hubrechts, L.; Huisman, J.A.; Jakeman, A.; et al. Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. *Adv. Water Resour.* 2009, 32, 147–158. [CrossRef]

8. Su, L.; Zhu, J.H.; Wang, W.; Liu, M. Application of ARMA Model in Prediction of Land Use Demand Take Farmland in Jin-Hu Coastal Area as Example. *Hunan Agric. Sci.* 2012, 5, 61–63.

9. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery. *Sensors* 2018, 18, 3717. [CrossRef]

10. Luus, F.P.; Salmon, B.P.; Bergh, F.V.D.; Maharaj, B.T. Multiview Deep Learning for Land-Use Classification. *IEEE Geosci. Remote Sens. Lett.* 2015, 12, 2448–2452. [CrossRef]
11. Mu, L.; Wang, L.; Wang, Y.; Chen, X.; Han, W. Urban Land Use and Land Cover Change Prediction via Self-Adaptive Cellular Based Deep Learning With Multisourced Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5233–5247. [CrossRef]
12. Azad, A.; Wang, X. Land Use Change Ontology and Traffic Prediction through Recurrent Neural Networks: A Case Study in Calgary, Canada. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 358. [CrossRef]
13. Kumar, S.; Radhakrishnan, N.; Mathew, S. Land use change modelling using a Markov model and remote sensing. *Geomat. Nat. Hazards Risk*** **2013**, *5*, 145–156. [CrossRef]
14. Lu, Y.; Wu, P.; Ma, X.; Li, X. Detection and prediction of land use/land cover change using spatiotemporal data fusion and the Cellular Automata–Markov model. *Environ. Monit. Assess.* **2019**, *191*, 68. [CrossRef] [PubMed]
15. Lu, C.-J.; Lee, T.-S.; Lian, C.-M. Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. *Decis. Support Syst.* **2012**, *54*, 584–596. [CrossRef]
16. Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C.; Villa-Vialaneix, N. Random Forests for Big Data. *Big Data Res.* **2017**, *9*, 28–46. [CrossRef]
17. Antoniadis, A.; Lambert-Lacroix, S.; Poggi, J.-M. Random forests for global sensitivity analysis: A selective review. *Reliab. Eng. Syst. Saf.* **2020**, *206*, 107312. [CrossRef]
18. Van de Geer, S.A. High-dimensional generalized linear models and the lasso. *Ann. Stat.* **2008**, *36*, 614–645. [CrossRef]
19. Wang, H.; Leng, C. Unified LASSO Estimation by Least Squares Approximation. *J. Am. Stat. Assoc.* **2007**, *102*, 1039–1048. [CrossRef]
20. Dyar, M.; Carmosino, M.; Breves, E.; Ozanne, M.; Clegg, S.; Wiens, R. Comparison of partial least squares and lasso regression techniques as applied to laser-induced breakdown spectroscopy of geological samples. *Spectrochim. Acta Part B At. Spectrosc.* **2012**, *70*, 51–67. [CrossRef]
21. Yilmaz, B.; Aras, E.; Nacar, S.; Kankal, M. Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. *Sci. Total Environ.* **2018**, *639*, 826–840. [CrossRef] [PubMed]
22. Bui, D.T.; Hoang, N.D.; Samui, P. Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (MARS). *J. Environ. Manag.* **2019**, *237*, 476–487. [CrossRef]
23. Nguyen, H.T.T.; Doan, T.M.; Radeloff, V. Applying random forest classification to map land use/land cover using landsat 8 oli. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, W4. [CrossRef]
24. Ha, T.V.; Tuothy, M.; Irwin, M.; Tuan, P.V. Monitoring and mapping rural urbanization and land use changes using Landsat data in the northeastern subtropical region of Vietnam. *Egypt. J. Remote Sens. Space Sci.* **2018**, *23*, 11–19. [CrossRef]
25. Dennedy-Frank, P.J.; Gorelick, S.M. Insights from watershed simulations around the world: Watershed service-based restoration does not significantly enhance streamflow. *Glob. Environ. Chang.* **2019**, *58*, 101938. [CrossRef]
26. Trinh-Tuan, L.; Matsumoto, J.; Ngo-Duc, T.; Hodzu, M.I.; Inoue, T. Evaluation of satellite precipitation products over Central Vietnam. *Prog. Earth Planet. Sci.* **2019**, *6*, 54. [CrossRef]
27. Luu, C.; Von Meding, J.; Kanjanabootra, S.; Pham, D. A proposed flood risk assessment method for Central Vietnam. In Proceedings of the 5th International Conference on Building Resilience, Newcastle, NSW, Australia, 15–17 July 2015; pp. 336–336–11. [CrossRef]
28. Van Khanh, N. Identify and Assess the Impact of Climate Change and Sea Level Rise to the System of Landfills and Solid Waste Treatment Facilities in the Central Coast Region of Vietnam. In *Waste Management and Resource Efficiency*; Springer: Singapore, 2018; pp. 195–208. [CrossRef]
29. Van Khanh, N. Identify and Assess the Impact of Climate Change and Sea Level Rise to the System of Landfills and Solid Waste Treatment Facilities in the Central Coast Region of Vietnam. In *Waste Management and Resource Efficiency*; Springer: Singapore, 2018; pp. 195–208. [CrossRef]
30. Overview of Quang Binh Province. Available online: https://www.quangbinh.gov.vn/web/en/detail?id=26029&_c=16407111 (accessed on 24 February 2022).
31. Quang Nam Portal. Available online: https://quangnam.gov.vn/webcenter/portal/ubnd_en (accessed on 24 February 2022).
32. Overview of Quang Tri Province. Available online: https://www.quangtri.gov.vn/xem-chi-tiet-gioi-thieu-tong-quan/-/view-article/1/3500113539863336577/157360322487QuangTri (accessed on 24 February 2022).
33. Population of Thua Thien Hue Province. Available online: https://thuathienhue.gov.vn/en-us/Home/Detail/tid/Population/ (accessed on 24 February 2022).
34. Overview of Quang Tri Province. Available online: https://www.quangtri.gov.vn/xem-chi-tiet-gioi-thieu-tong-quan/-/view-article/1/3500113539863336577/157360322487QuangTri (accessed on 24 February 2022).
35. Population of Thua Thien Hue Province. Available online: https://thuathienhue.gov.vn/en-us/Home/Detail/tid/Population/newsid/65f39533-85E0-4C1B-BC34-A8B600A82A8E/cid/AEBA5AE7-F4B9-4D9B-A507-DE8802BF1D14TTH (accessed on 24 February 2022).
36. Olyaie, E.; Ahyaneh, H.Z.; Mehr, A.D. A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. *Geosci. Front.* **2017**, *8*, 517–527. [CrossRef]
37. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
38. Sekulic, S.; Kowalski, B.R. MARS: A Tutorial. *J. Chemom.* **1992**, *6*, 199–216. [CrossRef]
39. Steinberg, D. An alternative to neural nets: Multivariate adaptive regression splines (MARS). *PC AI* **2001**, *15*, 38–41.
40. Fan, J.; Wu, L.; Ma, X.; Zhou, H.; Zhang, F. Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renew. Energy* **2020**, *145*, 2034–2045. [CrossRef]
41. LeBlanc, M.; Tibshirani, R. Adaptive principal surfaces. *J. Am. Stat. Assoc.* **1994**, *89*, 53–64. [CrossRef]
42. Sharda, V.N.; Prasher, S.O.; Patel, R.M.; Ojasvi, P.R.; Prakash, C. Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptatives (MARS) pour la prévision d’écoulement au sein de micro-bassins versants Himalayens d’altitudes intermédiaires avec peu de données. *Hydrol. Sci. J.* **2008**, *53*, 1165–1175.
41. Craven, P.; Wahba, G. Smoothing noisy data with spline functions. *Numer. Math.* 1978, 31, 377–403. [CrossRef]
42. Mohsen, S. Computational Estimation of Biliary Excretion of Compounds and the Role of Transporters. Ph.D. Thesis, University of Kent, Canterbury, UK, 2014.
43. Aydin, D.; Yilmaz, E. Modified Spline Regression based on Randomly Right-Censored Data: A Comparative Study. *Commun. Stat. Simul. Comput.* 2017, 47, 2587–2611. [CrossRef]
44. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 2010, 33, 1–22. [CrossRef]
45. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* 1996, 58, 267–288. [CrossRef]
46. Chouzenoux, E.; Pesquet, J.-C. A Stochastic Majorize-Minimize Subspace Algorithm for Online Penalized Least Squares Estimation. *IEEE Trans. Signal Process.* 2017, 65, 4770–4783. [CrossRef]
47. Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32. [CrossRef]
48. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* 2011, 66, 56–66. [CrossRef]
49. Rodríguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 2012, 67, 93–104. [CrossRef]
50. Rodríguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M.J.O.G.R. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 2015, 71, 804–818. [CrossRef]
51. Seni, G.; Elder, J.F. Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. *Synth. Lect. Data Min. Knowl. Discov.* 2010, 2, 1–126. [CrossRef]
52. Kardani, N.; Zhou, A.; Nazem, M.; Shen, S.-L. Estimation of Bearing Capacity of Piles in Cohesionless Soil Using Optimised Machine Learning Approaches. *Geotech. Geol. Eng.* 2019, 38, 2271–2291. [CrossRef]
53. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* 2018, 158, 1533–1543. [CrossRef]
54. Yang, J.-H.; Yang, M.-S. A control chart pattern recognition system using a statistical correlation coefficient method. *Comput. Ind. Eng.* 2005, 48, 205–221. [CrossRef]
55. Çevik, A.; Weber, G.-W.; Eyüboğlu, B.M.; Oğuz, K.K. Voxel-MARS: A method for early detection of Alzheimer’s disease by classification of structural brain MRI. *Ann. Oper. Res.* 2017, 258, 31–57. [CrossRef]
56. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* 2001, 106, 7183–7192. [CrossRef]
57. Ghorbani, M.A.; Deo, R.C.; Yaseen, Z.M.; Kashani, M.H.; Mohammadi, B. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. *Arch. Meteorol. Geophys. Bioclimatol. Ser. B* 2017, 133, 1119–1131. [CrossRef]
58. Jamali, A. Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: A case study of the northern region of Iran. *SN Appl. Sci.* 2019, 1, 1448. [CrossRef]
59. Adab, H.; Morbidelli, R.; Saltalippi, C.; Moradian, M.; Ghahari, G.A.F. Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data. *Water* 2020, 12, 3223. [CrossRef]
60. Duong, P.C.; Trung, T.H.; Nasahara, K.N.; Tadono, T. JAXA High-Resolution Land Use/Land Cover Map for Central Vietnam in 2007 and 2017. *Remote Sens.* 2018, 10, 1406. [CrossRef]