International Journal of Geo-Information

MDPI

*Article*

# Advancing Agricultural Land Suitability in Urbanized Semi-Arid Environments: Insights from Geospatial and Machine Learning Approaches

**Subbarayan Sathiyamurthi [1], Saravanan Subbarayan [2], Madhappan Ramya [3], Murugan Sivasakthi [3], Rengasamy Gobi [4], Saleh Qaysi [5], Sivakumar Praveen Kumar [3], Jinwook Lee [6], Nassir Alarifi [5], Mohamed Wahba [7] and Youssef M. Youssef [8,*]**

1 Horticultural Research Station, Tamil Nadu Agricultural University, Thadiyankudisai 624212, Tamil Nadu, India; sathya.soil@gmail.com
2 Department of Civil Engineering, National Institute of Technology (NITT), Tiruchirappalli 620015, Tamil Nadu, India; ssaravanan@nitt.edu
3 Department of Soil Science and Agriculture Chemistry, Faculty of Agriculture, Annamalai University, Annamalai Nagar 608002, Tamil Nadu, India; rammuagri21@gmail.com (M.R.); sivasarjai@gmail.com (M.S.); praveensk2099@gmail.com (S.P.K.)
4 Agricultural College and Research Institute, Karur 639001, Tamil Nadu, India; gobirenyagri@gmail.com
5 Department of Geology & Geophysics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia; sqaysi@ksu.edu.sa (S.Q.); nalarifi@ksu.edu.sa (N.A.)
6 Department of Civil, Environmental and Construction Engineering, University of Hawaii at Manoa, Honolulu, HI 96822, USA; jinwookl@hawaii.edu
7 Civil Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt; m_wahba@mans.edu.eg
8 Geological and Geophysical Engineering Department, Faculty of Petroleum and Mining Engineering, Suez University, Suez 43518, Egypt
* Correspondence: youssef.ibrahim@pme.suezuni.edu.eg

**Abstract:** Rising food demands are increasingly threatened by declining crop yields in urbanizing riverine regions of Southern Asia, exacerbated by erratic weather patterns. Optimizing agricultural land suitability (AgLS) offers a viable solution for sustainable agricultural productivity in such challenging environments. This study integrates remote sensing and field-based geospatial data with five machine learning (ML) algorithms—Naïve Bayes (NB), extra trees classifier (ETC), random forest (RF), K-nearest neighbors (KNN), and support vector machines (SVM)—alongside land-use/land-cover (LULC) considerations in the food-insecure Dharmapuri district, India. A grid searches optimized hyperparameters using factors such as slope, rainfall, temperature, texture, pH, electrical conductivity, organic carbon, available nitrogen, phosphorus, potassium, and calcium carbonate. The tuned ETC model showed the lowest root mean squared error (RMSE = 0.15), outperforming RF (RMSE = 0.18), NB (RMSE = 0.20), SVM (RMSE = 0.22), and KNN (RMSE = 0.23). The AgLS-ETC map identified 29.09% of the area as highly suitable (S1), 19.06% as moderately suitable (S2), 16.11% as marginally suitable (S3), 15.93% as currently unsuitable (N1), and 19.21% as permanently unsuitable (N2). By incorporating Landsat-8 derived LULC data to exclude forests, water bodies, and settlements, these suitability estimates were adjusted to 19.08% (S1), 14.45% (S2), 11.40% (S3), 10.48% (N1), and 9.58% (N2). Focusing on the ETC model, followed by land-use analysis, provides a robust framework for optimizing sustainable agricultural planning, ensuring the protection of ecological and social factors in developing countries.

**Keywords:** agricultural land suitability; machine learning; soil physicochemical; land-use/land-cover; Southern India

check for updates

**Citation:** Sathiyamurthi, S.; Subbarayan, S.; Ramya, M.; Sivasakthi, M.; Gobi, R.; Qaysi, S.; Praveen Kumar, S.; Lee, J.; Alarifi, N.; Wahba, M.; et al. Advancing Agricultural Land Suitability in Urbanized Semi-Arid Environments: Insights from Geospatial and Machine Learning Approaches. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 436. https://doi.org/10.3390/ijgi13120436

Academic Editors: Wolfgang Kainz, Dimitrios D. Alexakis and Adrianos Retalis

Received: 29 September 2024
Revised: 28 November 2024
Accepted: 30 November 2024
Published: 3 December 2024

**Copyright:** © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Urbanized riverine regions, with fertile soils and freshwater, are vital for agriculture and socioeconomic development [1–4]. However, the agricultural sector has faced increasing challenges in meeting food demands in developing countries over the past few decades [5,6], due to growing populations amidst climatic changes [7,8]. For instance, developing countries in South Asia (Figure 1a), with an anticipated number of inhabitants of approximately 2.1 billion by 2030, is projected to encounter substantial challenges in achieving sustainable agriculture and subsequently ensuring food demands to a rapidly increasing population [9–11] Another key challenge lies in the prevalence of traditional farming practices that often neglect crucial site suitability assessments [11]. This neglect leads to suboptimal crop yields and land degradation [12], further exacerbated by unsustainable practices that damage soil health, disrupt supply chains, and hinder agricultural production [3,13]. These stressors necessitate the identification of suitable agricultural lands to enhance food security, support sustainable production, and boost the carrying capacity of stressed riverine ecosystems [3,14,15].

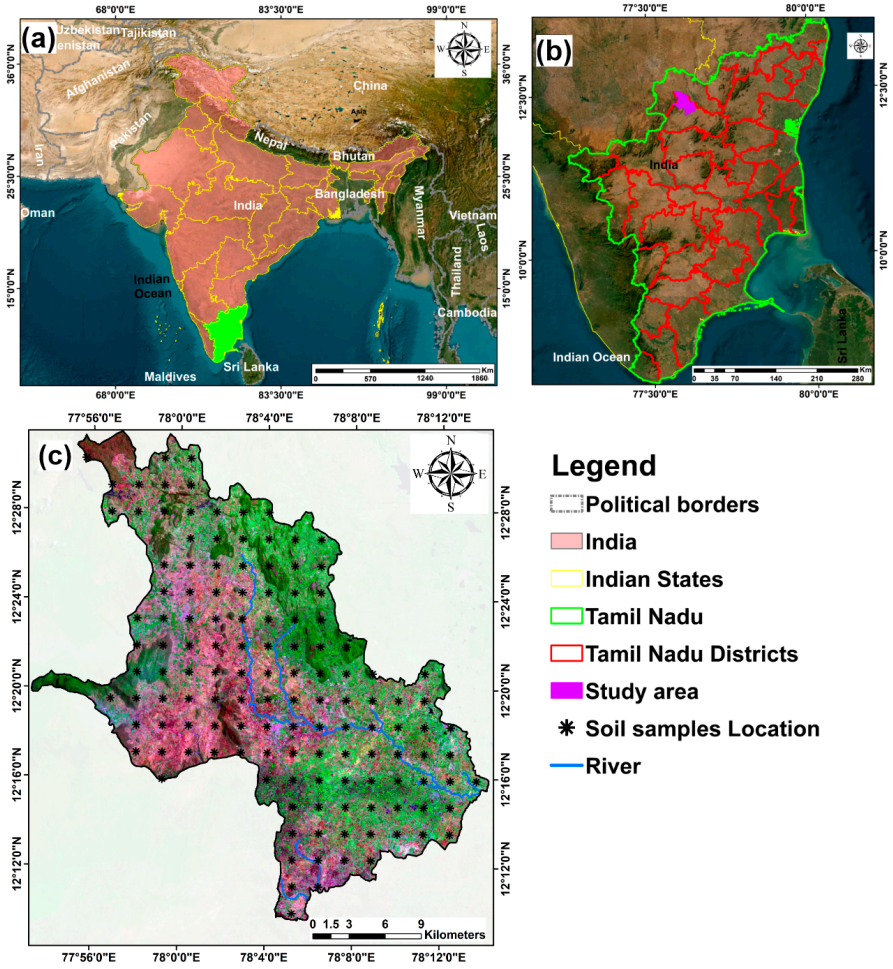


**Figure 1.** The location map represents Dharmapuri district within Tamil Nadu, India.

Consequently, agricultural land suitability (AgLS) analysis offers a structured geospatial technique specifically designed to identify the land parcels most suitable for agricultural activities [16,17]. This approach promotes efficient land resource utilization and minimizes environmental degradation by comprehensively evaluating the interplay between local environmental factors such as landscape characteristics, soil properties, and climatic conditions [6,18,19]. Numerous studies have identified key factors influencing AgLS, including topography, soil properties, climate conditions, and socioeconomic considerations across

various environments [3,17,20–22]. In this context, multispectral remote sensing has become an indispensable tool for mapping land-use/land-cover (LULC) changes in stressed regions [23], fostering a deeper understanding of the complex socioeconomic and environmental challenges [24,25].This comprehensive analysis offers a valuable tool for achieving several Sustainable Development Goals (SDGs), especially SDG 2 (Zero Hunger), SDG 11 (Sustainable Cities), and SDG 15 (Life on Land). This alignment is particularly crucial when considering peri-urban agriculture to gain traction within urbanized riverine environments [26,27], such as increasing food production to satisfy local demands, optimizing agricultural land-use practices, and ensuring food security [28,29].

Achieving a comprehensive assessment of site suitability requires a multidisciplinary approach that integrates surface and subsurface datasets [13]. In this regard, several geospatial land suitability analysis methods, including multi-criteria decision analysis (MCDA) [30], bivariate analysis [31], and machine learning (ML), have been widely developed for land suitability mapping [32–34]. AgLS mapping often employs a standardized suitability classification scheme (e.g., highly suitable, moderately suitable) to categorize land parcels based on their agricultural potential [35]. However, selecting the optimal method for AgLS mapping within geographic information system (GIS) environments remains a challenge, particularly in urbanized riverine areas in developing countries, due to varying data requirements for high-precision suitability maps [36]. For example, the analytical hierarchy process (AHP) is often criticized for its limitations in handling subjective decision-maker judgments [37]. These limitations, particularly ambiguity in factor importance, can significantly affect the accuracy of land suitability evaluations [38].

Alternatively, ML has emerged as a significant computational tool that can be applied in various aspects of ecosystems, including land suitability, yield forecasting, and crop pest estimation [39,40]. Unlike MCDA, ML can capture complex, non-linear relationships between influencing factors and suitability criteria [41–43] without requiring modifications to parameters, ratings, or weights within the site suitability framework [44,45]. Recent advancements in ML have led to the widespread application of various algorithms, including Naïve Bayes (NB), extra trees classifier (ETC), random forest (RF), K-nearest neighbors (KNN) [45], artificial neural networks (ANN) [46], support vector machines (SVM) [47], and bagging, for site suitability assessment [28,42,48]. For example, Taghizadeh-Mehrjardi et al. [41–43] applied ML to assess the suitability of rain-fed wheat and barley cultivation in Kurdistan, Iran, demonstrating superior accuracy and predictive performance compared to traditional methods. Agrawal et al. (2024) introduced an integrated framework combining MCDA and ML to evaluate AgLS for wheat and mustard in Abhanpur, Chhattisgarh [34]. Their Balanced Bagging Classifier achieved exceptional accuracies of 97.69% for mustard, 97.57% for wheat, and 97.36% for crop classification. Similarly, Khamoshi et al. (2024) [49] analyzed wheat cultivation suitability in Qazvin Province's Abiek region, where the random forest model achieved a higher accuracy (74%) compared to traditional approaches (62%), underscoring the precision and efficiency of ML in land suitability analysis. Building upon the success of ETC in land suitability mapping for combine harvesters [50], this study aims to further explore the potential of optimized ETC and other ML algorithms within the AgLS framework. This study bridges the knowledge gap by integrating SVM, KNN, RF, NB, and ETC models with remote sensing and field-based geospatial data, followed by LULC consideration. Addressing this research gap is crucial for developing data-driven techniques to tackle the pressing food insecurity issue in urbanized areas. The resulting optimized AgLS maps will be instrumental in guiding sustainable agricultural strategies to safeguard natural resources in these fragile ecosystems.

Agriculture is a cornerstone of the Indian economy, contributing nearly 40% of the nation's gross domestic product and serving as the primary source of income for approximately 70% of the population [48]. However, despite its significance, rain-fed agricultural practices dominate, covering nearly 67% of the country's cropland. The Green Revolution, a government initiative launched in the 1960s, was aimed at achieving self-sufficiency in food grain production [51]. The Dharmapuri district, located in Tamil Nadu state (Figure 1a–c),

was chosen as a case study due to its potential for cultivating high-value horticultural and floriculture crops (https://dharmapuri.nic.in/horticulture-and-plantation-crops/, accessed 23 February 2024). However, agricultural production in Dharmapuri faces challenges due to drought proneness and limited irrigation, as the region relies heavily on rain-fed practices [19]. Furthermore, Dharmapuri must balance agricultural needs with rapid urbanization and the expansion of intensive agriculture in the neighboring Salem district [52]. Studies conducted across India have demonstrated the effectiveness of ML techniques in AgLS [28,48,53]. Despite this potential, ML-based AgLS models have not yet been implemented in the Dharmapuri district. The implementation of such models could significantly contribute to optimizing land-use practices and enhancing agricultural productivity within this region.

This study addresses the challenge of improving agricultural land suitability assessments by evaluating the effectiveness of fine-tuned ML (SVM, KNN, RF, NB, and ETC) algorithms. The structured AgLS framework integrates fine-tuned ML models with remote sensing and field-based geospatial data, incorporating both surface (such as topography and climate) and subsoil physicochemical factors. This comparative approach aims to identify the optimal ML algorithm for generating the most precise AgLS map, considering land-use analysis in relation to environmental and social aspects in the Dharmapuri district. The approach employed has the potential to significantly enhance the long-term sustainability of agriculture in urbanized riverine ecosystems, benefiting communities globally.

## 2. Materials and Methods

### *2.1. Study Area Characteristics*

Dharmapuri district, situated in the northwestern corner of Tamil Nadu, lies at the confluence of Karnataka, Andhra Pradesh, and Tamil Nadu. Its geographical coordinates range from 11.47° and 12.53° North latitude and 77.28° to 78.45° East longitude. The research area encompasses six firkas—Marandahalli, Vellichandai, Palacode, Karimangalam, Periyanalli, and Pulikarai—spread across three taluks within Dharmapuri district. The elevation of the study area ranges from 127 to 1147 m above sea level (m a.s.l.). Over half of the region features a gentle slope, with gradients between 0% and 8%. The area supports a diverse range of crops, including mango, rice, paddy, ragi, pulses, and coconut, alongside exotic varieties such as dates (https://dharmapuri.nic.in/horticulture-and-plantation-crops/, accessed on 23 November 2024). The main water resource in the research region is the Cauvery River, flowing along its southwestern boundary. The area receives rainfall primarily from the southwest and northeast monsoons, with the northeast monsoon contributing the most. The average annual rainfall ranges from 650 to 900 mm. The maximum annual temperature, observed in April and May, reaches 38 °C, while the minimum temperature drops to 17 °C in December. Geologically, most of the study area is covered by granite gneiss interspersed with patches of schistose rocks, traversed by WNW—ESF and east—west trending dolerite dykes [54]. The region features a rocky, undulating terrain at its center, surrounded by hill ranges to the north and northwest (Mysore plateau), east (Javadu hill range), and south (Shevaroy hill ranges).

### *2.2. Research Methodology*

This study presents a holistic integration of space-borne data (including multispectral and topographic data) and soil physicochemical datasets through GIS-based ML models as a robust method to evaluate potential AgLS sites within the Dharmapuri district (Figure 2). The methodology, as shown in Figure 3, promotes the systematic integration of remote sensing data, field-based geospatial data, and GIS functionalities, as detailed below.

#### 2.2.1. Fieldwork and Data Preparation

Geospatial mapping methods were employed to highlight the spatial distribution and contributions of environmental covariates (topographic, climatic, and soil physicochemical), along with LULC. A GIS database for Dharmapuri was created from multiple datasets,

including Shuttle Radar Topography Mission (SRTM), Landsat-8 (L8), and the Climate Forecast System Reanalysis (CFSR) imagery, along with field soil sampling measurements. This data collection process was crucial for identifying potential suitable farming zones and ensuring the reliability and effectiveness of future analyses and decision-making processes.

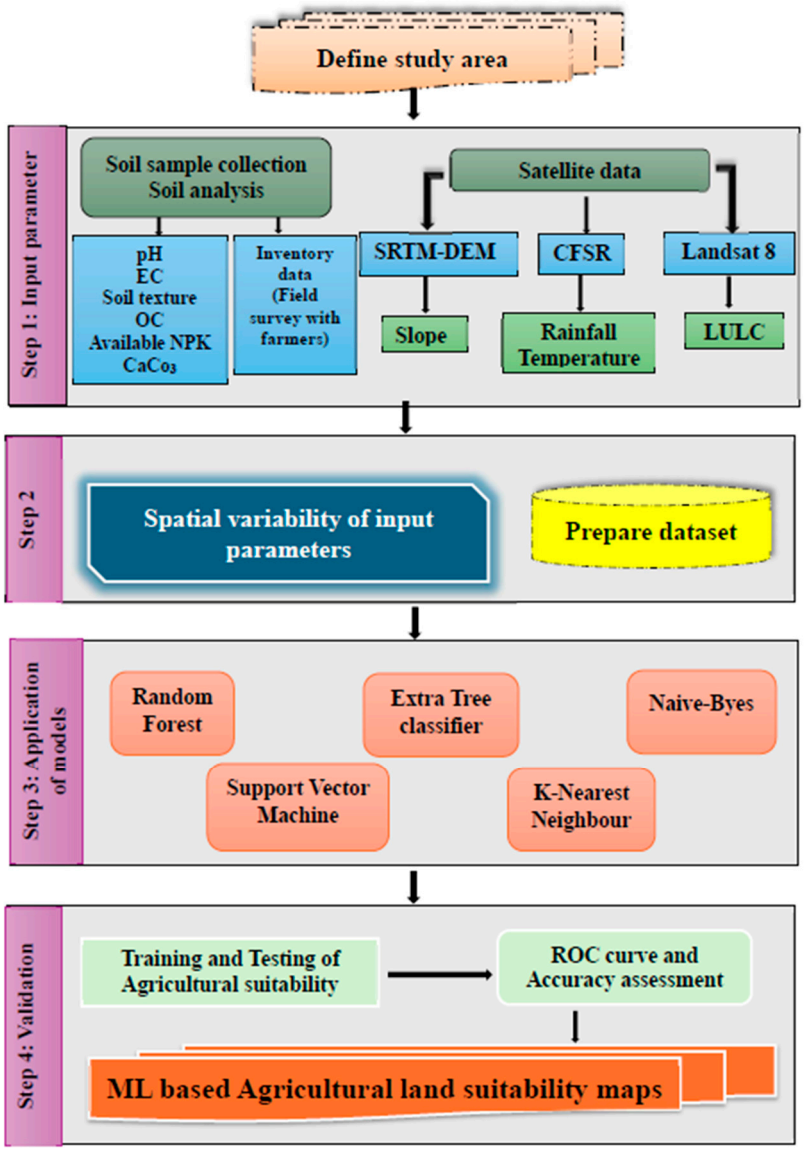


**Figure 2.** The methodology flow chart.

This study utilized daily precipitation and temperature data from 2000 to 2014, acquired from the global climate data via CFSR products (https://globalweather.tamu.edu/, accessed on 28 November 2024). The CFSR is a comprehensive, high-resolution reanalysis system covering atmospheric, oceanic, land surface, and sea ice domains designed to provide the best possible estimates of these coupled domains over the historical period [55]. Annual average climate datasets for temperature and rainfall were downloaded in CSV format for the entire study period and subsequently interpolated using the inverse distance weighting (IDW) method in ArcMap 10.8 (Figure 3a,b).

The Shuttle Radar Topography Mission-Digital Elevation Model (SRTM-DEM) data, obtained from the U.S. Geological Survey USGS EarthExplorer website (https://earthexplorer.usgs.gov/) (accessed on 12 May 2024), offers a high spatial resolution of 30 m (Figure 1). Subsequently, the spatial distribution of the surface slope was derived from a mosaic of the DEM image using ArcMap software (Figure 3c).

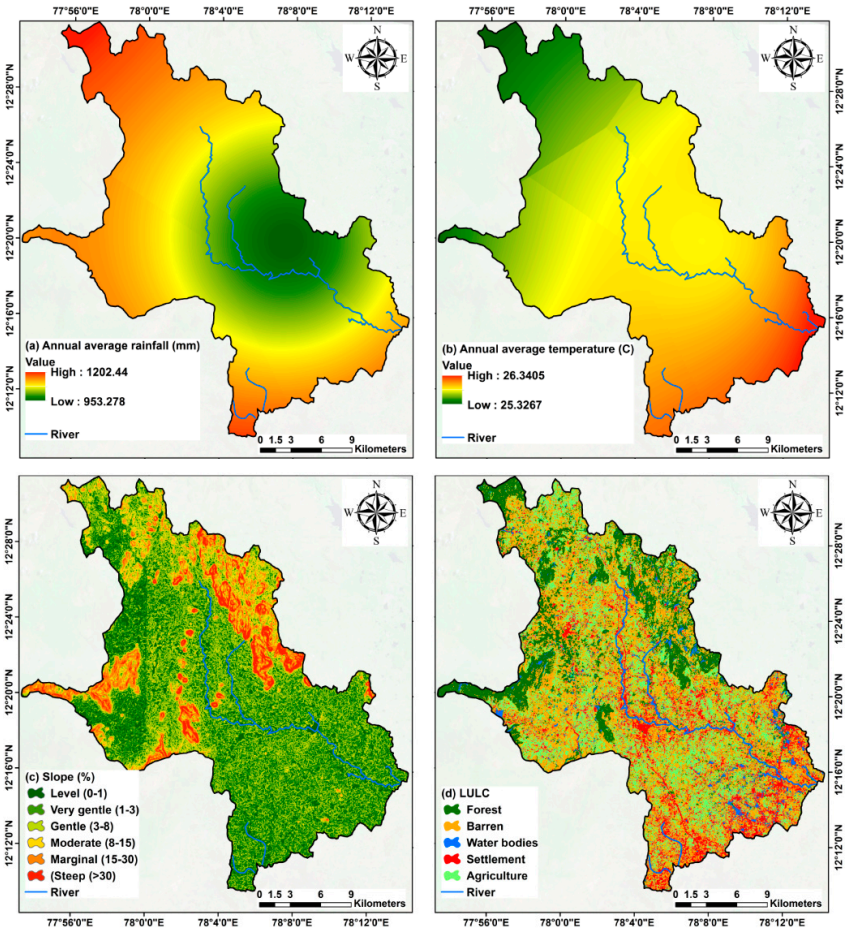


**Figure 3.** The controlling factors: (**a**) annual average rainfall; (**b**) annual average temperature; (**c**) slope; and (**d**) land use/land cover (LULC).

A LULC map for the study area was generated using multispectral satellite imagery acquired in 2021, aimed at identifying urbanization patterns and environmental features. Cloud-free scenes (less than 2% cloud cover) from paths 143 and 144, row 52, were retrieved from the USGS database (https://www.usgs.gov/landsat-missions, accessed on 15 March 2024). Non-thermal bands (3, 4, 5, 6, and 7) were calibrated to reflectance using radiometric and atmospheric correction tools available in the ENVI 5.1 software package. The Maximum Likelihood Classifier (MLC) was employed to create LULC maps, utilizing regions of interest (ROIs) derived from field observations and Google Earth Pro imagery. A random sample comprising 70% of the ROI database was used to train the classification model for the LULC maps. The accuracy of the classified images was validated by comparing 30% of the randomly selected ROIs against their corresponding classified pixels using a confusion matrix. The study area was then categorized into five LULC classes: forest, agriculture, water bodies, settlements, and barren lands (Figure 3c). The LULC maps demonstrate high reliability, achieving an overall accuracy of over 95% and a Kappa coefficient of 0.98.

This study leveraged a dataset of 125 soil samples collected within the Dharmapuri district by Ramya and Sathiyamurthi [56]. The samples were taken from the upper soil layer (0–15 cm) and included key physicochemical characteristics such as sand content (%), silt content (%), clay content (%), soil pH, electrical conductivity (EC; dS/m), soil organic carbon (SOC; g/kg), plant-available nitrogen (AN; kg/ha), available potassium (AK; kg/ha), and calcium carbonate (CaCO3; %). These samples were collected as both disturbed and undisturbed samples between February and April 2021 using a handheld Global Positioning System (GPS) device (GARMIN GPS Map 76CSx) across six selected firkas in the district (Figure 1). All the samples were stored in airtight containers following

collection for subsequent laboratory analysis based on the standard procedures (see Table 1). The collected soil data were converted into shapefile format using ArcGIS 10.4 software. Subsequently, spatial distribution maps for the soil pH, EC, soil texture, AN, AP, AK, and $CaCO_3$ content were generated using Kriging geostatistical interpolation methods (Figures 4–6). Descriptive statistics for the soil variables were computed using the data analysis tools in the Microsoft Excel software package (Table 2).

**Table 1.** Standard method for analysis of soil physicochemical properties.

| S. No. | Parameter | Methodology Followed | Reference |
|---|---|---|---|
| 1 | Soil reaction (pH) | 1: 2.5 soil water suspension with pH meter | Jackson [57] |
| 2 | Electrical conductivity ($dSm^{-1}$) | 1: 2.5 soil water suspension with electrical conductivity bridge | Jackson [57] |
| 3 | Textural analysis | International pipette method | Piper [58] |
| 4 | Soil organic carbon (%) | Chromic acid wet digestion method | Walkley and Black [59] |
| 5 | Available N ($Kg\ ha^{-1}$) | Alkaline Permanganate method | Subbiah and Asija (63) |
| 6 | Available P ($Kg\ ha^{-1}$) | Colorimetric method using 0.5 M Na $HCO_3$ (pH 8.5) | Olsen et al. [60] |
| 7 | Available K($Kg\ ha^{-1}$) | Flame photometer method | Standford and English [61] |
| 8 | $CaCO_3$ content (%) | Rapid titration method | Piper [58] |

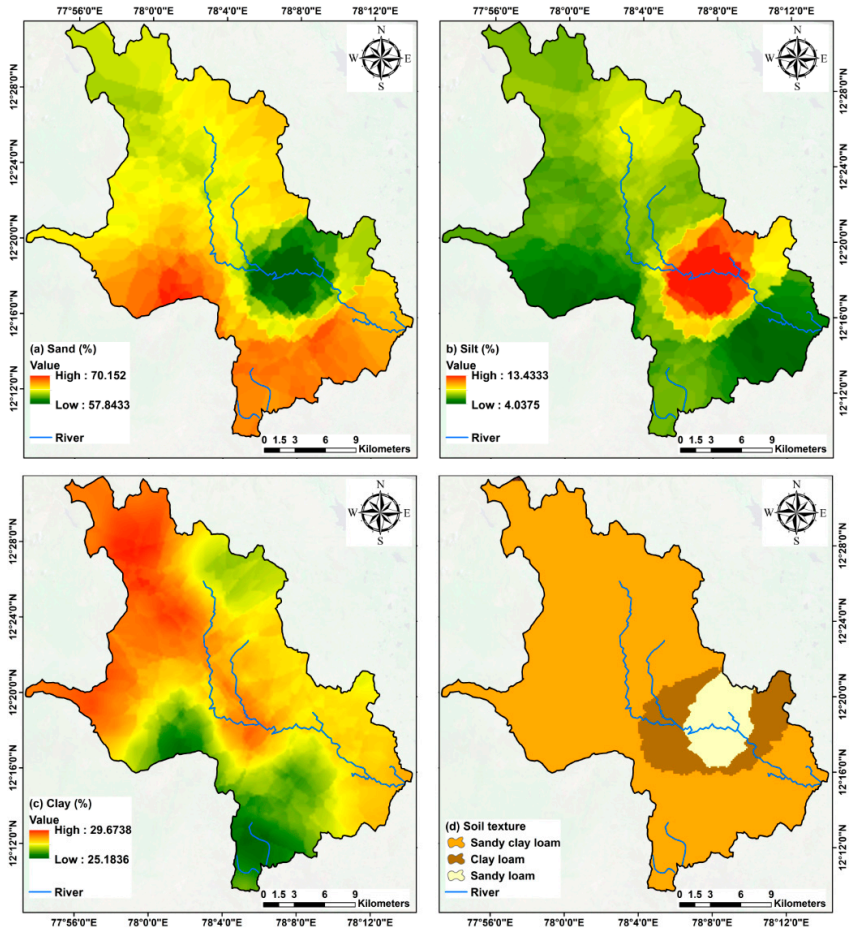


**Figure 4.** The controlling factors: (**a**) sand (%); (**b**) silt (%); (**c**) clay (%); and (**d**) soil texture.

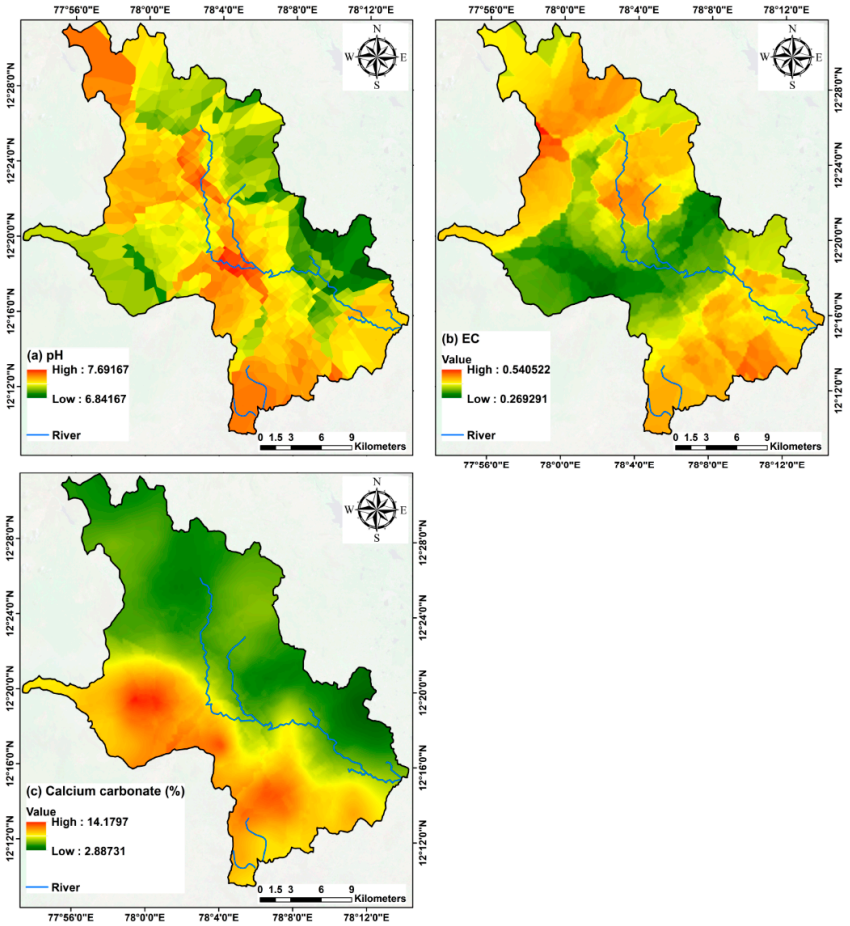


**Figure 5.** The controlling factors: (**a**) pH; (**b**) EC; and (**c**) calcium carbonate content (%).

**Table 2.** Descriptive statistics of the soil parameters in the study area.

| Soil Properties | Min * | Max | Mean | SD * | CV * | Kurtosis | Skewness |
|---|---|---|---|---|---|---|---|
| pH | 6 | 8.3 | 7.4 | 0.44 | 0.04 | −1.20 | 2.01 |
| EC dS m-1 | 0.14 | 1.30 | 0.34 | 0.18 | 0.02 | 2.52 | 8.62 |
| Sand (%) | 41.5 | 73.8 | 65.8 | 4.75 | 0.52 | −2.41 | 10.01 |
| Silt (%) | 2 | 29.3 | 5.75 | 3.98 | 0.44 | 4.09 | 20.35 |
| Clay (%) | 21.7 | 32.2 | 27.7 | 2.17 | 0.24 | −0.36 | 0.17 |
| Soil organic carbon | 0.2 | 13.6 | 5.2 | 3.12 | 0.34 | 0.51 | −0.42 |
| Available nitrogen (Kg $ha^{-1}$) | 116 | 332 | 223 | 45.39 | 5.04 | −0.007 | −0.05 |
| Available phosphorus (Kg $ha^{-1}$) | 3 | 77 | 33 | 14.60 | 1.62 | 0.35 | −0.38 |
| Available potassium (Kg $ha^{-1}$) | 65 | 517 | 225 | 106.03 | 11.78 | 0.81 | 0.20 |
| Calcium carbonate (%) | 1 | 30 | 6.5 | 4.69 | 0.52 | 1.48 | 3.59 |

* Min—Minimum, Max—Maximum, SD—Standard deviation, CV—coefficient of Variation.

#### 2.2.2. Land Suitability Assessment Using Various Machine Learning Models

ML is a subfield of Artificial Intelligence (AI) focused on the development of predictive models from complex and potentially unstructured datasets [62,63]. This study investigated the utility of various ML algorithms in generating a highly accurate AgLS map for the Dharmapuri district, incorporating land-use analysis that reflects both environmental and social considerations. Eleven critical covariates influencing AgLS were identified and included in the analysis: slope; annual average temperature (AT); annual average rainfall (AR); soil texture (sand, silt, and clay content); soil pH; EC; SOC; AN; AP; AK; and $CaCO_3$ content. All the variables were resampled to a uniform spatial resolution of 30 m using raster analysis tools in ArcMap software. Furthermore, the thematic layers were coregistered to a Universal Transverse Mercator (UTM) projection (WGS84, zone

44N) to ensure spatial alignment. A group of 200 farmers was selected from a group of 500, based on a minimum landholding criterion of 2 acres. The participants comprised educated and progressive farmers capable of understanding and implementing recommendations from the Department of Agriculture, Tamil Nadu, India. The evaluation focused on assessing soil fertility across two suitability categories: suitable (high productivity) and unsuitable (degraded land, saline soils, and waterlogged areas). These categories were used as inventory data and validated through analysis using Google Earth Pro. The field survey was conducted with the cooperation of landowners familiar with their fields, ensuring accurate data collection on the perceived soil conditions. The sampling strategy followed the guidelines established by the Tamil Nadu Agricultural University (https://tnau.ac.in/site/kvk-salem/soil-and-water-testing-3/, accessed 12 May 2023). Subsequently, five optimized ML algorithms—SVM, RF, KNN, XGBoost, and NB—were employed to develop predictive models for AgLS classification using Python software. Hyperparameter optimization was conducted using GridSearchCV and 10-fold cross-validation to minimize training errors and enhance model generalization [64]. The resulting AgLS map was reclassified using the natural breaks method within ArcMap (Figures 7–11), categorizing land parcels into suitability classes (Table 3), namely highly suitable ($S_1$), moderately suitable ($S_2$), marginally suitable ($S_3$), currently not suitable ($N_1$), and permanently not suitable ($N_2$). Following AgLS map generation, the land-use/land-cover (LULC) map was used to extract and exclude critical features from the analysis, including forest areas, water bodies, and settlements (Table 4). This step refines the AgLS map by incorporating social and environmental considerations specific to the Dharmapuri district. A detailed explanation of each employed ML algorithm is provided in the following section.

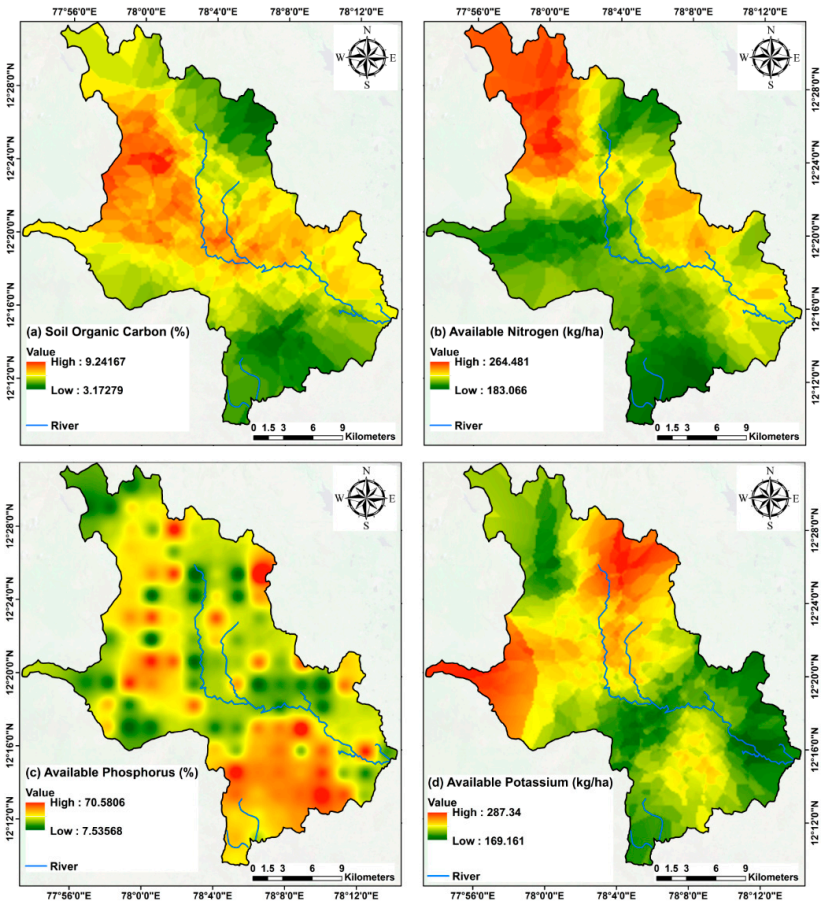


**Figure 6.** The controlling factors: (**a**) soil organic carbon (%); (**b**) available nitrogen; (**c**) available phosphorus (%); and (**d**) available potassium.

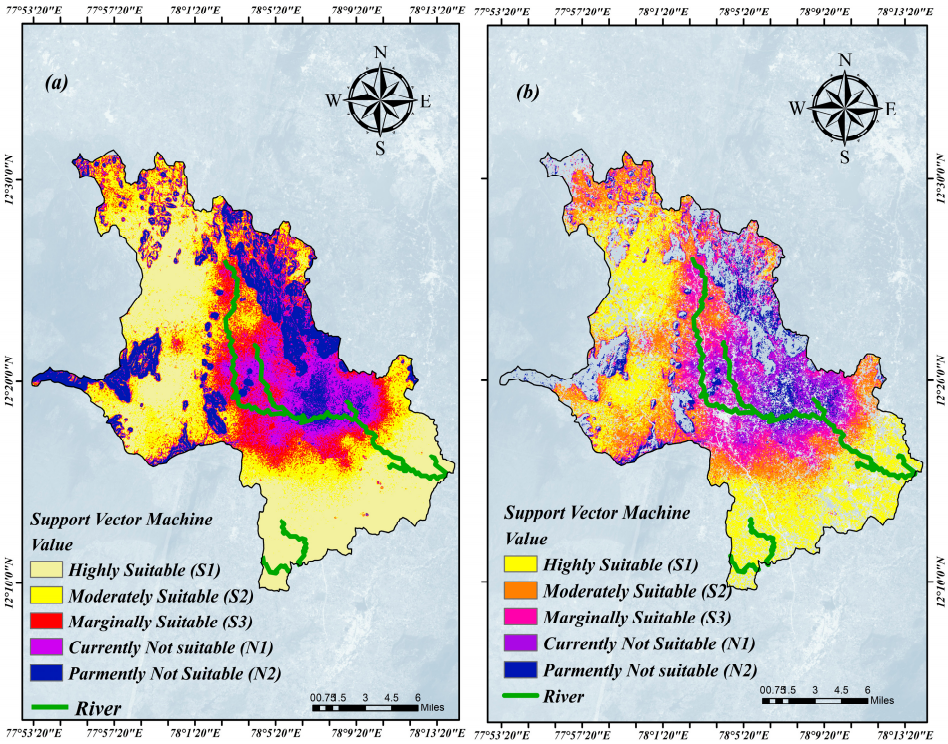


**Figure 7.** Agriculture land suitability (AgLS) map based on SVM algorithm before (**a**) and after (**b**) mask removal.

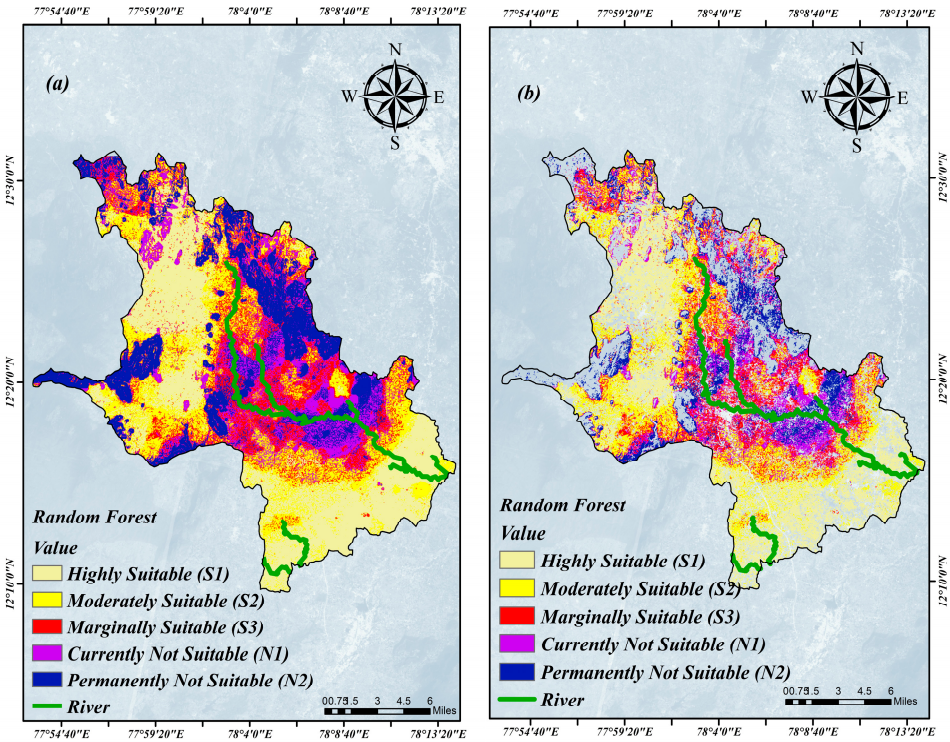


**Figure 8.** Agriculture land suitability (AgLS) map based on random forest (RF) algorithm before (**a**) and after (**b**) mask removal.

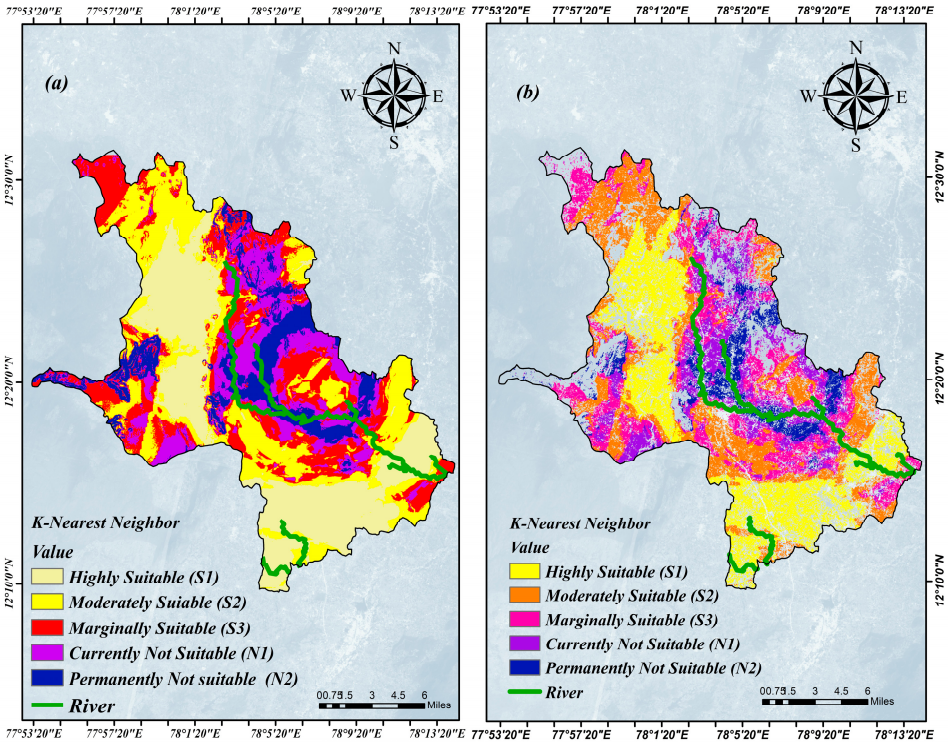


**Figure 9.** Agriculture land suitability (AgLS) map based on K-nearest neighbor (KNN) algorithm before (**a**) and after (**b**) mask removal.

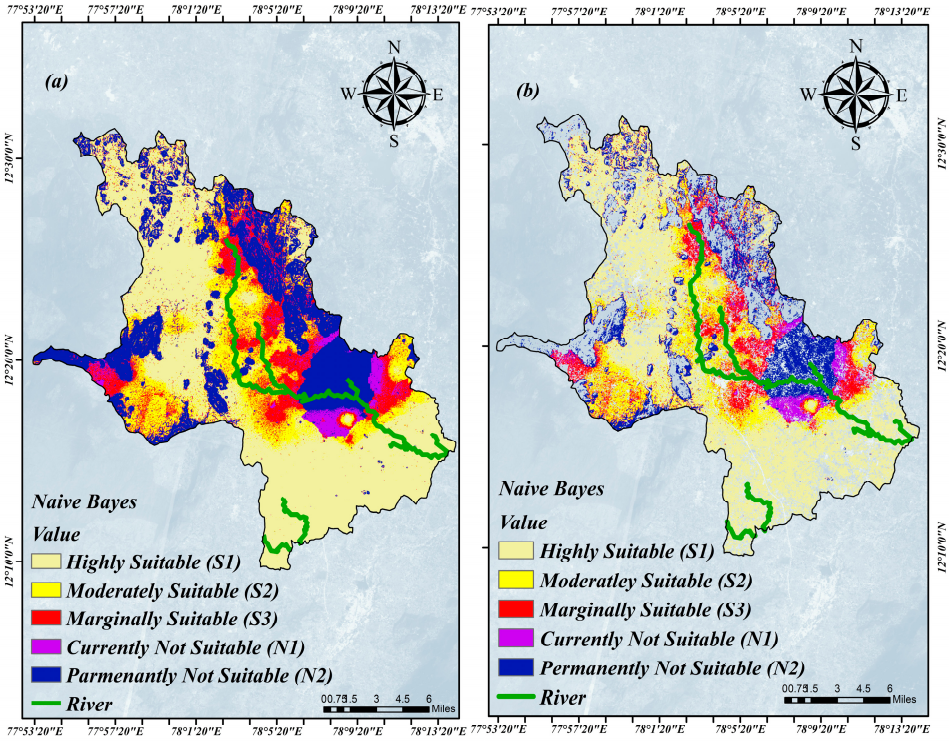


**Figure 10.** Agriculture land suitability (AgLS) map based on Naïve Bayes (NB) algorithm before (**a**) and after (**b**) mask removal.

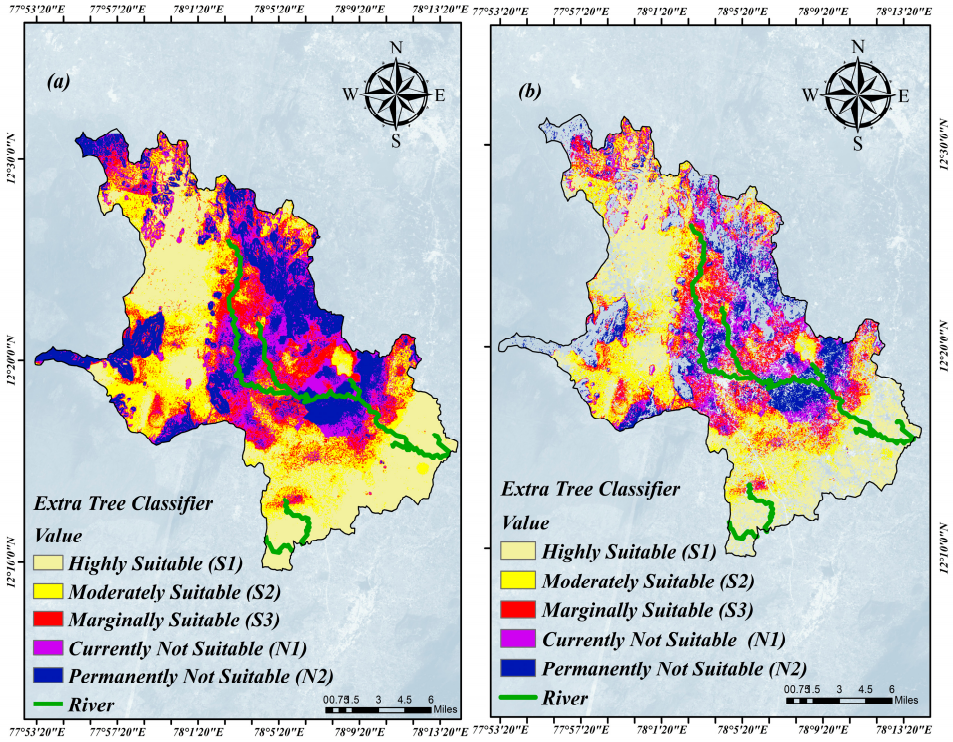


**Figure 11.** Agriculture land suitability (AgLS) map based on Extra Tree Classifier (ETC) algorithm before (**a**) and after (**b**) mask removal.

**Table 3.** ML-based land suitability classes.

| Suitability Class | SVM | | RF | | KNN | | NB | | ETC | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) |
| Highly suitable ($S_1$) | 228.52 | 35.45 | 195.06 | 30.24 | 177.07 | 27.45 | 278.41 | 43.26 | 187.60 | 29.09 |
| Moderately suitable ($S_2$) | 128.69 | 19.95 | 133.25 | 20.65 | 161.90 | 25.10 | 107.93 | 16.75 | 126.80 | 19.06 |
| Marginally suitable ($S_3$) | 97.65 | 15.14 | 99.90 | 15.50 | 132.09 | 20.49 | 72.92 | 11.35 | 103.91 | 16.11 |
| Currently not suitable ($N_1$) | 97.04 | 15.04 | 98.99 | 15.35 | 101.48 | 15.74 | 33.97 | 5.28 | 102.73 | 15.93 |
| Permanently not suitable ($N_2$) | 92.88 | 14.42 | 117.71 | 18.26 | 72.36 | 11.22 | 150.67 | 23.36 | 123.87 | 19.21 |

SVM—Support vector machine; RF—Random forest; KNN—k-nearest neighbor; NB—Naïve Bayes; ETC—Extra tree classifier.

**Table 4.** ML-based land suitability classes after masking.

| Suitability Class | SVM | | RF | | KNN | | NB | | ETC | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) | Area ($km^2$) | Area (%) |
| Highly suitable ($S_1$) | 153.07 | 23.73 | 129.75 | 20.13 | 118.95 | 18.45 | 190.32 | 29.52 | 123.05 | 19.08 |
| Moderately suitable ($S_2$) | 94.31 | 14.62 | 97.54 | 15.12 | 112.68 | 17.40 | 77.58 | 12.02 | 93.16 | 14.45 |
| Marginally suitable($S_3$) | 67.27 | 10.43 | 70.31 | 10.90 | 85.83 | 13.70 | 50.78 | 7.87 | 73.50 | 11.40 |
| Currently not suitable ($N_1$) | 64.99 | 10.45 | 65.14 | 10.10 | 61.73 | 9.58 | 24.23 | 4.28 | 67.60 | 10.48 |
| Permanently not suitable ($N_2$) | 37.14 | 5.76 | 54.04 | 8.38 | 37.61 | 5.86 | 72.89 | 11.30 | 59.48 | 9.58 |

SVM—Support vector machine; RF—Random Forest; KNN—k-Nearest Neighbor; NB—Naïve Bayes; ETC—Extra tree classifier.

### Support Vector Machine Model (SVM)

SVM represents a supervised learning method widely employed for both classification and regression tasks. Notably, SVM has the capability to handle non-linear relationships within the data [65]. The core principle of the SVM algorithm lies in generating a hyperplane, essentially a multidimensional line, that effectively separates data points belonging

to distinct categories. This separation is achieved by identifying a small subset of data points closest to the hyperplane, known as support vectors. These support vectors play a crucial role in defining the optimal position and orientation of the hyperplane, thereby maximizing the margin between the classified categories. A key strength of SVM lies in its ability to effectively categorize data points even in scenarios where the underlying data exhibit non-linear characteristics. In essence, SVM functions as supervised learning algorithms that leverage training data to establish a hyperplane capable of differentiating between distinct classes [66]. This classification strategy prioritizes maximizing the separation distance between categories relative to the hyperplane, which ultimately minimizes the influence of noise and enhances model generalizability [67].

In a training set of $n$ data points, $Y_i$ can take on the values $-1$ or $+1$, indicating which class each point $X_i$ belongs to. Where $\|W\|$ is the norm of the normal of the hyperplane, $b$ denotes the $n$-dimensional hyperplane that maximally separates the two classes shown as

$$Y_i((W * X_i) + b) \geq 1 \quad (1)$$

The scalar base is usually denoted by $b$. By employing Equation (3), the Lagrangian formula is presented and applied to the situation. The objective is to find the value of the Lagrangian multiplier in restriction, which minimizes the Lagrangian $L$ concerning $\|W\|$ and $b$.

$$L = 1/2\|W\|^2 - \sum_{i=1}^{n} \lambda_i(Y_i((W* X_i) + b) - 1) \quad (2)$$

The kernel approaches have solid theoretical underpinnings and practical efficacy. When the kernel function $K(x_i, x)$ is used, the SVM model's decision boundary may be written as:

$$f(x) = \sum_{i=1}^{L} y_i\lambda_i K(x_i, x) + b \quad (3)$$

The SVM model excels in handling unfamiliar data structures due to its kernel technique, which simplifies the resolution of various problems. Selecting the kernel function for SVM models is critical yet complex. In this study, a linear kernel was employed and the cost (C), epsilon ($\varepsilon$), and gamma ($\gamma$) parameters optimized through trial-and-error to enhance model performance [68].

Random Forest Model (RF)

Random forest (RF) is an ensemble learning technique that utilizes a multitude of decision trees to enhance classification or regression tasks [69]. Each tree in the forest is built using a random subset of features from the original data, promoting diversity and reducing overfitting. During tree construction, a random subset of features is considered for splitting at each node, further increasing the robustness of the model. Two key parameters for RF optimization are the number of trees in the forest and the minimum number of samples required for a terminal node [70]. The out-of-bag (OOB) error, derived from observations not used in tree training (unchecked datasets), serves as a valuable tool for assessing model performance. Random forests demonstrably enhance classification accuracy and reduce the overfitting risk by incorporating an ensemble of decision trees [70]. Additionally, RF models can handle non-normalized data due to their reliance on tree-based rules. However, building and training a large number of trees can be computationally intensive and time-consuming. Furthermore, the interpretability of RF models can be challenging due to their ensemble nature, and assessing the relative importance of individual variables requires specialized techniques. The performance of the RF model was enhanced through the tuning of several crucial parameters (Table 5), including the number of decision trees (n_estimators), maximum depth of each decision tree in the forest (max_depth), and the lowest number of samples essential to divide an internal node (min_samples_split) or a leaf node (min_samples_leaf).

**Table 5.** Performance metrics of the ML-based AgLS models, along with the tuning parameter of ML algorithms.

| Model/Algorithm | Cohen-Kappa (κ) | AUC | RMSE | Tuning Parameter |
|---|---|---|---|---|
| Support vector machine | 0.60 | 0.86 | 0.22 | Kernel = Linear, c = 200, ε = 0.5, γ = Scale |
| Random forest | 0.68 | 0.89 | 0.18 | n_estimator = 100, max_depth = none, Min_samples_split = 2, Min_sample_leaf = 1, and Max_features = auto |
| K-nearest neighbor (KNN) | 0.53 | 0.81 | 0.23 | n_neighbors = 5, weights = uniform, algorithm = brute, metric = minkowski, and verbose = 1 |
| Naïve Bayes | 0.60 | 0.83 | 0.20 | Alpha = 1.4, Fit_prior = false, and Class_prior = none |
| Extra tree classifier | 0.69 | 0.89 | 0.15 | Bootstrap = False, criterion = Gini, min_samples_leaf = 1, min_samples_split = 2, and n_estimators = 2080 |

A large number of unrelated decision tree models are built to complete the random forest training process. The final outcome of the classification process is the mode of the data, and each of these distinctive decision trees offers a prediction about the mode of the sample. By adding more training sets that are unrelated to one another, the effectiveness of the random forest model is increased.

The output of the random forest, based on the many classifications learned from training sets, is determined by

$$H(x) = \text{argmax}_z \sum_{i=1}^{k} I(h(x) = Z) \quad (4)$$

$Z$ is the outcome variable, and $I(\cdot)$ is the indicative function. In this case, the single decision tree model $h(i)$ replaces the RF model $H(x)$.

K-Nearest Neighbors Algorithm

K-nearest neighbors (KNN) is one of the simplest and non-parametric ML algorithms used for classification and regression problems, and it performs well with large training sets. It is based on the closest k-distance, i.e., the Euclidean distance, in the data sets (plurality or mean). KNN has worked based on the feature similarity of k, the closest distance in the training example in the data sets [71]. The significance of the k-NN algorithm is that it does not build any model, unlike other ML algorithms, and does not require any training dataset separately. KNN model performance was optimized through hyperparameter tuning, including the number of neighbors, weight assignment, algorithm selection, distance metric, and verbose output settings [72]. The nearest neighbor methods computed the predicted value of their k-nearest neighbors and all the training data sets. The K neighbors are selected using the Euclidean distance. The Euclidean distance is the most common method used in KNN to calculate the distance between two data points. The formula for Euclidean distance is as follows:

$$d(x, y) = \sqrt{(x_1 - x_2)^2 - (y_1 - y_2)^2} \quad (5)$$

The algorithms can be used to create maps, estimate small areas, and make inferences without requiring any specific assumptions about the distribution of predictor variables [73].

Naïve Bayes (NB) Algorithm

Naïve Bayes (NB) is a classification algorithm based on Bayes' theorem, a fundamental concept in probability theory. A key advantage of NB is its computational efficiency, as it requires less training data compared to more complex algorithms. This makes NB particularly well-suited for situations with limited datasets. However, the NB model assumes independence between features, which may not always hold true in real-world applications. Despite this limitation, NB can be a powerful tool for various classification

tasks, including document classification, spam filtering, and more. Several variants of NB exist, including Gaussian, Multinomial, Complement, Bernoulli, and Categorical, each suited for different data types [74]. The multinomial Naïve Bayes model was optimized through hyperparameter tuning by adjusting the Laplace smoothing parameter (alpha), prior probability estimation (fit_prior), and class prior probability values (Table 5).

Extra Trees Classifier Algorithm

The extra trees classifier (ETC) is an ensemble machine learning algorithm that leverages decision trees for classification tasks. ETC shares similarities with random forests but incorporates key distinctions. Unlike random forests, which rely on bootstrapped subsets of the training data to build individual trees, ETC utilizes the entire training dataset for each tree construction. While both algorithms introduce randomness during tree building, ETC employs random thresholds at each split point within a tree. In contrast, random forests consider a random subset of features at each node. This approach fosters the creation of a forest comprising decorrelated decision trees, consequently enhancing the overall robustness of the model. Similar to random forests, final predictions in ETC are derived by aggregating the results (e.g., majority vote) from all the trees within the ensemble [75]. The Gini index and entropy can be calculated using the following equation:

$$\text{Gini Impurity} = \sum_{j=1}^{O} f_j(1 - f_j) \tag{6}$$

$$\text{Entropy} = -\sum_{j=1}^{O} f_j \log(f_j) \tag{7}$$

where $f_i$ is the number of unique class labels and $f_j$ is the proportion of rows with output label $i$. To optimize ETC model performance, hyperparameter tuning was conducted focusing on bootstrap sampling (Bootstrap), splitting criteria(criterion), minimum leaf and split sample sizes, and the number of trees (Table 5).

2.2.3. Model Performance Assessment, Comparison, and Validation

Model validation is a crucial step in land suitability analysis, ensuring the generated AgLS map accurately reflects real-world conditions [76]. This study employed various ML techniques to assess the performance of the AgLS maps on a dedicated test dataset. A confusion matrix was used to evaluate model performance by categorizing predictions into true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) [77]. An analysis of the confusion matrix provides valuable insights into the model's effectiveness and classification accuracy [44,45].

Three key metrics—area under the curve (AUC), root mean squared error (RMSE), and Cohen-kappa (κ)—were employed to assess the model's ability to predict agricultural land suitability during hyperparameter trial-and-error tuning (Table 5). RMSE measures the magnitude of the difference between predicted and actual values, with lower RMSE values indicating better model performance. AUC, derived from the receiver operating characteristic (ROC) curve, ranges from 0.1 to 1.0, with higher values indicating superior model performance [77].

$$\text{AUC} = \frac{\varepsilon TP + \varepsilon TN}{(P + N)} \tag{8}$$

where $P$ is the total number of lands for agricultural production and $N$ is the total number of lands not suitable for agricultural production.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} (\rho_i - \sigma_i)^2}{m}} \tag{9}$$

where $\rho_i$ = prediction, $\sigma_i$ = actual value, $z$ = the mean of actual values, $m$ = total count of data.

## 3. Results and Discussion

### 3.1. Geospatial Variability of Environmental Covariates

Understanding spatial variability of soil properties is critical for optimizing agricultural management and ensuring food security, even within individual soil layers [78]. Table 2 summarizes the descriptive statistics for the soil physicochemical properties across the study area. The analysis revealed a range of mean values for the investigated properties, from 0.34 to 225. Notably, sand, clay, AN, AK, and AP exhibited higher mean values (>27.7), while the remaining characteristics displayed lower means (<7.4). According to [79], the soil property variability was assessed using the coefficient of variation (CV) as low ($CV \leq 15\%$), medium ($15\% < CV \leq 35\%$), and high ($CV > 35\%$). The CV results indicated low variability for all the soil properties, with values ranging from 0.02% to 11.78%. Soil chemical content (AK, AN, and AP) displayed the greatest variation and standard deviations, followed by sand, $CaCO_3$, silt, SOC, and clay, respectively. This suggests a wider spread in the data for the chemical elements compared to textural fractions. Kurtosis values for all soil properties ranged from −0.007 to 4.09 (Table 2). Furthermore, the analysis revealed substantial skewness in pH, EC, sand, silt, $CaCO_3$, and AK, indicating deviations from a normal distribution [80]. The observed high variability in soil characteristics likely arises from a complex interplay of inherent soil properties under agricultural management practices and prevailing climatic conditions [81]. The slope is a well-recognized critical topographic factor influencing agricultural land suitability, with the study area exhibiting a range of slopes, including gentle slopes of less than 8% grade (Figure 3c). Sand reached its maximum in the southern parts (Figure 4a). The high content of clay (29.7%) is found in the northwestern part, while silt (13.4%) prevails as a local high in the middle part (Figure 4b,c). The superficial texture map displays three soil types: sandy clay loam, sandy loam, and clay loam (Figure 4d). Regions with higher rainfall are typically more favorable for agricultural activities, while areas with lower rainfall are less conducive to such practices [82] Within the study area, the annual average rainfall ranges from 953 mm to 1202 mm (Figure 3a). The northeastern region receives relatively low rainfall (<900 mm), whereas higher rainfall levels are observed in the northern, northwestern, and western parts of the area. The spatial distribution map of average annual temperatures indicates that higher temperatures are concentrated in the southern regions (Figure 3b), while lower temperatures (<25.32 °C) are predominantly found in the northwestern part of the area.

### 3.2. Mapping Agricultural Land Suitability Using Machine Learning Models

This study introduces a comprehensive framework for assessing land suitability for sustainable agriculture, integrating topographical, climatic, and soil physicochemical factors, along with LULC consideration. Five tuned ML methods—SVM, RF, KNN, NB, and ETC—were utilized in this study. Eleven input variables were used for model development, encompassing one topographic factor (slope), two climatic factors (AT and AR), and eight soil properties (soil texture, pH, EC, SOC, AN, AK, AP, and $CaCO_3$). Optimal hyperparameters for each model were determined (Table 5) and subsequently applied to generate AgLS maps across the study area. A classification scheme was implemented to categorize the maps into five suitability classes based on probability values, with higher values indicating greater agricultural potential (Figures 7–11).

The findings indicate that the proportions of highly suitable ($S_1$) agricultural lands identified by ML algorithms are 35.45% for SVM, 30.24% for RF, 27.45% for KNN, 43.26% for Naïve Bayes, and 29.09% for ETC. Areas deemed moderately suitable ($S_2$) are 19.95% for SVM, 20.65% for RF, 25.1% for KNN, 16.75% for Naïve Bayes, and 19.06% for ETC. The marginally suitable ($S_3$) areas are 15.14% for SVM, 15.5% for RF, 20.49% for KNN, 11.35% for Naïve Bayes, and 16.11% for ETC. Areas classified as currently unsuitable ($N_1$) are 15.04% for SVM, 15.35% for RF, 15.74% for KNN, 5.28% for Naïve Bayes, and 15.93% for ETC. The permanently not suitable ($N_2$) areas are 14.42% for SVM, 18.26% for RF, 11.22% for KNN, 23.36% for Naïve Bayes, and 19.21% for ETC.

To refine these results, satellite-derived LULC data, highlighting forests, water bodies, and settlements, were used to mask these features from the AgLS models. Consequently, the revised highly agricultural suitability zones for Dharmapuri district are 23.73% for SVM, 20.13% for RF, 18.45% for KNN, 29.52% for Naïve Bayes, and 19.08% for ETC. Moderately suitable areas are 14.62% for SVM, 15.12% for RF, 17.4% for KNN, 12.02% for Naïve Bayes, and 14.45% for ETC. Marginally suitable areas are 10.43% for SVM, 10.9% for RF, 13.7% for KNN, 7.87% for Naïve Bayes, and 16.11% for ETC. Currently, unsuitable areas are 10.45% for SVM, 10.1% for RF, 9.58% for KNN, 4.28% for Naïve Bayes, and 10.48% for ETC. Permanently unsuitable areas are 5.76% for SVM, 8.38% for RF, 5.86% for KNN, 11.30% for Naïve Bayes, and 9.58% for ETC.

The results reveal that each ML model produces varying classifications for the suitability zones and their spatial distribution. Notable differences were observed in the distribution of the N2 and N1 classes, which are predominantly located in the northern and northeastern parts of the study area. In contrast, highly and moderately suitable zones are primarily concentrated in the southeastern and western regions of the district. These findings highlight the spatial variability and unique predictive capabilities of each ML model in assessing agricultural land suitability. Although these models will identify optimal zones for agricultural expansion in Dharmapuri, it is crucial to conduct an initial evaluation of the predictive model's accuracy before applying AgLS maps in planning [83]. The AUC of the ROC analysis, derived from the model's testing dataset, can be categorized as excellent (0.9–1.0), very good (0.8–0.9), good (0.7–0.8), average (0.6–0.7), and poor (0.5–0.6). As highlighted by [84], ML offers a significant advantage by eliminating the requirement for expert knowledge in traditional rule-based approaches. Consequently, all employed models exhibited very good precision in the agricultural land suitability assessment (Figure 12). However, a more comprehensive statistical evaluation revealed that the ETC model (AUC = 0.89, κ = 0.69, and RMSE = 0.15) outperformed the RF (AUC = 0.89, κ = 0.68, and RMSE = 0.18), SVM (AUC = 0.86, κ = 0.60, and RMSE = 0.22), NB (AUC = 0.83, κ = 0.60, and RMSE = 0.20), and KNN (AUC = 0.81, κ = 0.53, and RMSE = 0.23) models (Table 5).

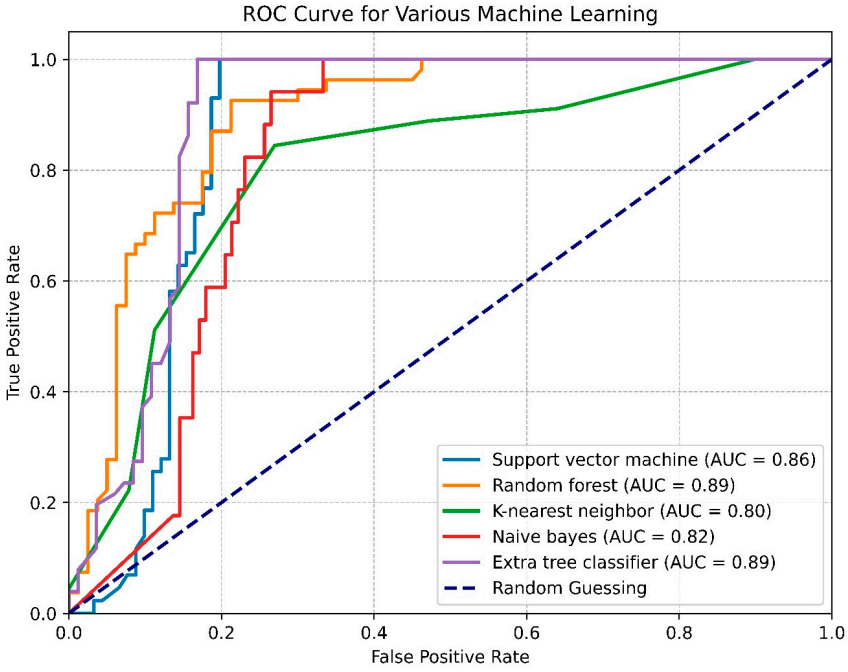


**Figure 12.** ROC curve of all machine learning algorithms.

The findings of this study are consistent with recent research in similar Asian environments, as documented by [42,48,53,85–87]. For instance, refs. [48,85] reported superior speed and accuracy of RF compared to SVM in eye state estimation. In Iran, ref. [42]

also utilized RF and SVM algorithms to create land suitability maps for barley and wheat. Ref. [53] further employed an RF model in Darjeeling (India) to predict land suitability for tea cultivation, achieving an impressive AUC accuracy of 85.2%. Ref. [86] used SVM and ANN to recommend suitable crops based on land-quality parameters. Ref. [87] proposed an ML-based approach for evaluating tea cultivation suitability in Anhui province, China. RF, a well-established ensemble learning method, is renowned for its high accuracy [88], achieved through the construction of numerous decision trees using randomly selected features. However, this study demonstrated that the extra trees classifier (ETC) model achieved a lower RMSE (0.15) and outperformed RF in certain metrics. Unlike RF, which generates decision trees independently, ETC leverages multiple, decorrelated decision trees, enhancing accuracy and evaluation metrics [89]. The ETC model's superior performance in this study can be attributed to its robust hyperparameter tuning, which mitigates overfitting and optimizes weight adjustments, as discussed by Probst et al. (2019) [90]. Despite its strengths, the ETC model's performance varies depending on the application. For instance, ref. [50] reported a lower accuracy (AUC of 60%) for ETC when predicting land suitability for combine harvesters. In contrast, RF's performance often suffers from limited hyperparameter tuning, resulting in less effective weight adjustments compared to ETC [91]. Thus, while both RF and ETC have proven effective, their relative performance is context-dependent, influenced by model configuration and dataset characteristics.

Consequently, the ETC model emerges as the most suitable selection for AgLS analysis within an urbanized riverine territory, assisting in developing cost-effective, socially acceptable, and environmentally sustainable strategies. Analysis of the optimal AgLS-ETC map generated using the employed ETC algorithms revealed that 29.09%, 19.06%, 16.11%, 15.93%, and 19.21% of the investigated area fell within the classifications of highly suitable ($S_1$), moderately suitable ($S_2$), marginally suitable ($S_3$), currently unsuitable ($N_1$), and permanently not suitable ($N_2$) zones (Figure 11a and Table 3), respectively. Conversely, when forests, water bodies, and settlements were excluded from the analysis, the corresponding proportions of the total area categorized as highly, moderately, marginally, currently, and permanently not suitable zones were 19.08%, 14.45%, 11.40%, 10.48%, and 9.58% (Figure 11b and Table 4), respectively. Corroborating previous research by Aguilar-Rivera et al. [92] on the dependence of crop yield suitability on climatic changes, Kilic et al. [82] further suggest a positive association between high rainfall and areas highly suitable for agriculture. This alignment supports the delineation of the highly suitable zones ($S_1$) observed in the southeastern and western regions of the ETC map (Figure 11). This spatial distribution aligns with the topographical and climatic variations across the district, indicating that these factors, in conjunction with soil properties, play a crucial role in determining agricultural suitability. Furthermore, the integration of LULC data ensures that the identified suitable areas are realistic and applicable, avoiding overestimation by excluding non-agricultural land uses. This comprehensive and multi-faceted approach provides a robust tool for land-use planners and policymakers aiming to promote sustainable agricultural practices, optimize land resources, and enhance food security in the region. In this context, this study provides valuable insights for farmers, supporting sustainable agricultural practices and enhancing crop yield stability in comparable environmental conditions.

### *3.3. Study Limitations and Future Works*

Efficient utilization of agriculturally productive land for sustainable practices is vital to ensuring long-term food security. Accurate land suitability modeling, which accounts for the spatial variability of soil properties, can significantly boost agricultural productivity in urbanized riverine regions of tropical developing countries. This study presents a comprehensive framework for evaluating land suitability for sustainable agriculture by integrating detailed analyses of soil physicochemical properties, geomorphological characteristics, and climatic conditions, employing five optimized ML algorithms within these riverine ecosystems. However, several limitations exist when applying ML-based land suitability analyses. For instance, the choice of predictor variables, tailored to specific environmental

settings, can profoundly influence the assessment's outcomes. Incorporating additional factors, such as LULC, often enhances model performance, making the careful selection of variables critical. Moreover, the robustness of ML models depends heavily on the availability of extensive ground-based soil and inventory observations. As highlighted by Carvalho et al. [93], large datasets are essential to ensure effective training and validation phases, thus mitigating issues with generalizability. Incorporating socioeconomic, political, and cultural factors alongside biophysical parameters could further improve model accuracy. Future research should also integrate crop-specific suitability analyses using ground-truth data for various crops, including mango, rice, paddy, ragi, pulses, and coconut [94]. This approach will enhance the precision and reliability of the generated models, aligning land-use strategies more effectively with local agricultural priorities.

## 4. Conclusions

Global food security is threatened by declining crop yields in urbanizing riverine regions with increasingly erratic weather patterns, particularly in semi-arid areas of developing Asian countries. To address this issue, this study integrates remote sensing and field-based geospatial data with five tuned ML algorithms (NB, ETC, RF, KNN, SVM) to assess AgLS in Dharmapuri district (India), incorporating topographic, climatic, and subsoil physicochemical factors. The findings reveal that 187.6 $km^2$ (ETC), 278.41 $km^2$ (NB), 177.07 $km^2$ (KNN), 195.06 $km^2$ (RF), and 228.52 $km^2$ (SVM) of land are highly suitable for cultivation. LULC data derived from satellite imagery were used to highlight forests, water bodies, and settlements, facilitating the creation of a mask to exclude these features from the AgLS models. Consequently, the revised highly suitable areas, excluding forests, water bodies, and settlements, are 123.05 $km^2$ (ETC), 190.32 $km^2$ (NB), 118.95 $km^2$ (KNN), 129.75 $km^2$ (RF), and 153.07 $km^2$ (SVM) for the respective models. The optimized ETC model demonstrated superior performance (AUC = 0.89, Kappa = 0.69, RMSE = 0.15) when compared to the other algorithms. These findings underscore the novel application of the ETC algorithm as an excellent framework for agrarian land suitability, directly aiding farmers in selecting alternate suitable lands. This study demonstrates the potential of machine learning-based AgLS modeling to enhance agricultural sustainability through optimized land-use planning, while mitigating environmental and social impacts in urbanizing riverine regions.

**Author Contributions:** Conceptualization, Subbarayan Sathiyamurthi and Youssef M. Youssef; Data curation, Subbarayan Sathiyamurthi and Madhappan Ramya; Formal analysis, Saravanan Subbarayan, Saleh Qaysi, Sivakumar Praveen kumar, Jinwook Lee and Nassir Alarifi; Funding acquisition, Nassir Alarifi and Youssef M. Youssef; Investigation, Subbarayan Sathiyamurthi and Saravanan Subbarayan; Methodology, Subbarayan Sathiyamurthi, Madhappan Ramya, Rengasamy Gobi and Youssef M. Youssef; Project administration, Subbarayan Sathiyamurthi; Resources, Murugan Sivasakthi, Rengasamy Gobi and Sivakumar Praveen kumar; Software, Murugan Sivasakthi, Rengasamy Gobi, Sivakumar Praveen kumar and Mohamed Wahba; Supervision, Subbarayan Sathiyamurthi and Youssef M. Youssef; Validation, Saravanan Subbarayan, Saleh Qaysi, Jinwook Lee, Nassir Alarifi, Mohamed Wahba and Youssef M. Youssef; Visualization, Subbarayan Sathiyamurthi and Youssef M. Youssef; Writing—original draft, Subbarayan Sathiyamurthi and Youssef M. Youssef; Writing—review and editing, Saleh Qaysi, Jinwook Lee, Nassir Alarifi, Mohamed Wahba and Youssef M. Youssef. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Researchers Supporting Project number (RSP2024R432), King Saud University, Riyadh, Saudi Arabia.

**Data Availability Statement:** The original soil raw data can be found at https://data.mendeley.com/datasets/nm5b82wbmt/3, accessed on 28 November 2024.

**Acknowledgments:** The authors would like to express their gratitude to the Researchers Supporting Project number (RSP2024R432) at King Saud University in Riyadh, Saudi Arabia. We are also indebted to the participating farmers for their valuable time and knowledge in sharing their understanding of soil fertility. Finally, the authors thank Tamilnadu Agricultural University for the soil collection protocol and Annamalai University for its unwavering support with the soil analysis.

**Conflicts of Interest:** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## References

1. Heino, M.; Kinnunen, P.; Anderson, W.; Ray, D.K.; Puma, M.J.; Varis, O.; Siebert, S.; Kummu, M. Increased probability of hot and dry weather extremes during the growing season threatens global crop yields. *Sci. Rep.* **2023**, *13*, 3583. [CrossRef] [PubMed]
2. Gopi, P.; Karthikeyan, M. Intelligent Crop Recommendation with Yield Prediction using Dragonfly Algorithm based Deep Learning Model. In Proceedings of the 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2–4 February 2023; IEEE: New York, NY, USA, 2023; pp. 880–885. [CrossRef]
3. Arafa, N.A.; El-Said Salem, Z.; Abdeldayem, A.L.; Ghorab, M.A.; Moustafa, Y.M.; Soliman, S.A.; Farag, M.H.; Purohit, S.; Elhag, M.; Youssef, Y.M. Advancing Deltaic Aquifer Vulnerability Mapping to Seawater Intrusion and Human Impacts in Eastern Nile Delta: Insights from Machine Learning and Hydrochemical Perspective. *Earth Syst. Environ.* **2024**, 1–26. [CrossRef]
4. Sathiyamurthi, S.; Sivasakthi, M.; Saravanan, S.; Gobi, R.; Kumar, S.P.; Karuppannan, S. Assessment of crop suitability analysis using AHP-TOPSIS and geospatial techniques: A case study of Krishnagiri District, India. *Environ. Sustain. Indic.* **2024**, *24*, 100466. [CrossRef]
5. Reynolds, T.W.; Waddington, S.R.; Anderson, C.L.; Chew, A.; True, Z.; Cullen, A. Environmental impacts and constraints associated with the production of major food crops in Sub-Saharan Africa and South Asia. *Food Secur.* **2015**, *7*, 795–822. [CrossRef]
6. FAO. *The State of Food Security and Nutrition in the World 2022*; FAO: Rome, Italy, 2022. [CrossRef]
7. Bhanumathi, S.; Vineeth, M.; Rohit, N. Crop Yield Prediction and Efficient use of Fertilizers. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 4–6 April 2019; IEEE: New York, NY, USA, 2019; pp. 0769–0773. [CrossRef]
8. Gupta, S.; Chatterjee, P.; Rastogi, R.; Gonzalez, E.D.S. A Delphi fuzzy analytic hierarchy process framework for criteria classification and prioritization in food supply chains under uncertainty. *Decis. Anal. J.* **2023**, *7*, 100217. [CrossRef]
9. Funk, C.C.; Brown, M.E. Declining global per capita agricultural production and warming oceans threaten food security. *Food Secur.* **2009**, *1*, 271–289. [CrossRef]
10. Funk, C.; Dettinger, M.D.; Michaelsen, J.C.; Verdin, J.P.; Brown, M.E.; Barlow, M.; Hoell, A. Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11081–11086. [CrossRef]
11. Islam, R.; Ahmed, R.; Dey, B.; Haque, S.; Aktar, S.; Bhuiyan, S.; Arif, M.S.; Ador, A.H.; Haque, M.M.U.; Saha, N. Salinity hazard drives the alteration of occupation, land use and ecosystem service in the coastal areas: Evidence from the south-western coastal region of Bangladesh. *Heliyon* **2023**, *9*, e18512. [CrossRef]
12. Ghanem, S.; Rukia, A. Distribution of soil organic matter in the coastal region of Syria: A case study. *DYSONA-Appl. Sci.* **2020**, *1*, 57–63.
13. Tavakoli, M.; Motlagh, Z.K.; Sayadi, M.H.; Ibraheem, I.M.; Youssef, Y.M. Sustainable Groundwater Management Using Machine Learning-Based DRASTIC Model in Rurbanizing Riverine Region: A Case Study of Kerman Province, Iran. *Water* **2024**, *16*, 2748. [CrossRef]
14. Mendas, A.; Delali, A. Integration of MultiCriteria Decision Analysis in GIS to develop land suitability for agriculture: Application to durum wheat cultivation in the region of Mleta in Algeria. *Comput. Electron. Agric.* **2012**, *83*, 117–126. [CrossRef]
15. Liu, R.; Zhang, K.; Zhang, Z.; Borthwick, A.G. Land-use suitability analysis for urban development in Beijing. *J. Environ. Manag.* **2014**, *145*, 170–179. [CrossRef] [PubMed]
16. Bonfante, A.; Monaco, E.; Langella, G.; Mercogliano, P.; Bucchignani, E.; Manna, P.; Terribile, F. A dynamic viticultural zoning to explore the resilience of terroir concept under climate change. *Sci. Total Environ.* **2018**, *624*, 294–308. [CrossRef] [PubMed]
17. Hussain, S.; Nasim, W.; Mubeen, M.; Fahad, S.; Tariq, A.; Karuppannan, S.; Alqadhi, S.; Mallick, J.; Almohamad, H.; Abdo, H.G. Agricultural land suitability analysis of Southern Punjab, Pakistan using analytical hierarchy process (AHP) and multi-criteria decision analysis (MCDA) techniques. *Cogent Food Agric.* **2024**, *10*, 2294540. [CrossRef]
18. Song, X.-P.; Huang, W.; Hansen, M.C.; Potapov, P. An evaluation of Landsat, Sentinel-2, Sentinel-1 and MODIS data for crop type mapping. *Sci. Remote Sens.* **2021**, *3*, 100018. [CrossRef]
19. Datta, P.; Behera, B.; Rahut, D.B. Climate change and Indian agriculture: A systematic review of farmers' perception, adaptation, and transformation. *Environ. Chall.* **2022**, *8*, 100543. [CrossRef]
20. Rathi, R.; Khanduja, D.; Sharma, S. Synergy of fuzzy AHP and Six Sigma for capacity waste management in Indian automotive industry. *Decis. Sci. Lett.* **2015**, *4*, 441–452. [CrossRef]

21. Sathiyamurthi, S.; Saravanan, S.; Sankriti, R.; Aluru, M.; Sivaranjani, S.; Srivel, R. Integrated GIS and AHP techniques for land suitability assessment of cotton crop in Perambalur District, South India. *Int. J. Syst. Assur. Eng. Manag.* **2024**, *15*, 267–278. [CrossRef]
22. Islam, M.M.; Ahamed, T.; Matsushita, S.; Noguchi, R. A Damage-Based Crop Insurance System for Flash Flooding: A Satellite Remote Sensing and Econometric Approach. In *Remote Sensing Application II. New Frontiers in Regional Science: Asian Perspectives*; Ahamed, T., Ed.; Springer: Singapore, 2024; Volume 77. [CrossRef]
23. Singh, A. Review Article Digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **1989**, *10*, 989–1003. [CrossRef]
24. Shalaby, A.; Tateishi, R. Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Appl. Geogr.* **2007**, *27*, 28–41. [CrossRef]
25. Khalil, M.M.E.; Khodary, S.M.; Youssef, Y.M.; Alsubaie, M.S.; Sallam, A. Geo-Environmental Hazard Assessment of Archaeological Sites and Archaeological Domes—Fatimid Tombs—Aswan, Egypt. *Buildings* **2022**, *12*, 2175. [CrossRef]
26. Duan, W.; Chen, Y.; Zou, S.; Nover, D. Managing the water-climate- food nexus for sustainable development in Turkmenistan. *J. Clean. Prod.* **2019**, *220*, 212–224. [CrossRef]
27. Wanyama, D.; Mighty, M.; Sim, S.; Koti, F. A spatial assessment of land suitability for maize farming in Kenya. *Geocarto Int.* **2021**, *36*, 1378–1395. [CrossRef]
28. Dey, B.; Ferdous, J.; Ahmed, R. Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables. *Heliyon* **2024**, *10*, e25112. [CrossRef] [PubMed]
29. Tian, X.; Engel, B.A.; Qian, H.; Hua, E.; Sun, S.; Wang, Y. Will reaching the maximum achievable yield potential meet future global food demand? *J. Clean. Prod.* **2021**, *294*, 126285. [CrossRef]
30. Elsadek, W.M.; Wahba, M.; Al-Arifi, N.; Kanae, S.; El-Rawy, M. Scrutinizing the performance of GIS-based analytical Hierarchical process approach and frequency ratio model in flood prediction—Case study of Kakegawa, Japan. *Ain Shams Eng. J.* **2023**, *15*, 102453. [CrossRef]
31. Wahba, M.; Hassan, H.S.; Elsadek, W.M.; Kanae, S.; Sharaan, M. Prediction of Flood Susceptibility Using Frequency Ratio Method: A Case Study of Fifth District, Egypt. 2022. Available online: https://www.iche2022.org/_files/ugd/21d103_be56d464954b4310a2461ccf02fd83b9.pdf (accessed on 25 August 2024).
32. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [CrossRef]
33. Yin, S.; Li, J.; Liang, J.; Jia, K.; Yang, Z.; Wang, Y. Optimization of the Weighted Linear Combination Method for Agricultural Land Suitability Evaluation Considering Current Land Use and Regional Differences. *Sustainability* **2020**, *12*, 10134. [CrossRef]
34. Agrawal, N.; Govil, H.; Kumar, T. Agricultural land suitability classification and crop suggestion using machine learning and spatial multicriteria decision analysis in semi-arid ecosystem. *Environ. Dev. Sustain.* **2024**, 1–38. [CrossRef]
35. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2020.
36. Romano, G.; Sasso, P.D.; Liuzzi, G.T.; Gentile, F. Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy. *Land Use Policy* **2015**, *48*, 131–143. [CrossRef]
37. Akıncı, H.; Özalp, A.Y.; Turgut, B. Agricultural land use suitability analysis using GIS and AHP technique. *Comput. Electron. Agric.* **2013**, *97*, 71–82. [CrossRef]
38. Xue, L.; Cao, P.; Xu, D.; Guo, Y.; Wang, Q.; Zheng, X.; Han, R.; You, A. Agricultural land suitability analysis for an integrated rice–crayfish culture using a fuzzy AHP and GIS in central China. *Ecol. Indic.* **2023**, *148*, 109837. [CrossRef]
39. Ransom, C.J.; Kitchen, N.R.; Camberato, J.J.; Carter, P.R.; Ferguson, R.B.; Fernández, F.G.; Franzen, D.W.; Laboski, C.A.; Myers, D.B.; Nafziger, E.D.; et al. Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. *Comput. Electron. Agric.* **2019**, *164*, 104872. [CrossRef]
40. Arrouays, D.; Poggio, L.; Guerrero, O.A.S.; Mulder, V.L. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. *Geoderma Reg.* **2020**, *21*, e00265. [CrossRef]
41. Wang, Z.; Lai, C.; Chen, X.; Yang, B.; Zhao, S.; Bai, X. Flood hazard risk assessment model based on random forest. *J. Hydrol.* **2015**, *527*, 1130–1141. [CrossRef]
42. Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Rasoli, L.; Kerry, R.; Scholten, T. Land Suitability Assessment and Agricultural Production Sustainability Using Machine Learning Models. *Agronomy* **2020**, *10*, 573. [CrossRef]
43. Roell, Y.E.; Beucher, A.; Møller, P.G.; Greve, M.B.; Greve, M.H. Comparing a Random Forest Based Prediction of Winter Wheat Yield to Historical Yield Potential. *Agronomy* **2020**, *10*, 395. [CrossRef]
44. Møller, A.B.; Mulder, V.L.; Heuvelink, G.B.M.; Jacobsen, N.M.; Greve, M.H. Can We Use Machine Learning for Agricultural Land Suitability Assessment? *Agronomy* **2021**, *11*, 703. [CrossRef]
45. El-Rawy, M.; Wahba, M.; Fathi, H.; Alshehri, F.; Abdalla, F.; El Attar, R.M. Assessment of groundwater quality in arid regions utilizing principal component analysis, GIS, and machine learning techniques. *Mar. Pollut. Bull.* **2024**, *205*, 116645. [CrossRef]
46. Wahba, M.; Sharaan, M.; Elsadek, W.M.; Kanae, S.; Hassan, H.S. Examination of the efficacy of machine learning approaches in the generation of flood susceptibility maps. *Environ. Earth Sci.* **2024**, *83*, 429. [CrossRef]
47. Wahba, M.; Sharaan, M.; Elsadek, W.M.; Kanae, S.; Hassan, H.S. Building Information Modeling Integrated with Environmental Flood Hazard to Assess the Building Vulnerability to Flash Floods. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 1–21. [CrossRef]

48. Sarkar, D.; Saha, S.; Maitra, M.; Mondal, P. Site suitability for Aromatic Rice cultivation by integrating Geo-spatial and Machine learning algorithms in Kaliyaganj C.D. block, India. *Artif. Intell. Geosci.* **2021**, *2*, 179–191. [CrossRef]
49. Khamoshi, S.E.; Sarmadian, F.; Omid, M. Land suitability evaluation using traditional and machine learning approaches: A case study in abiek plain, Qazvin province, Iran. *Iran. J. Soil Water Res.* **2024**, *55*, 269–283.
50. Rahman, M.; Kamruzzaman, M.; Shahid, S.; Thorp, K.R.; Rahaman, H.; Shahriyar, M.; Islam, A.K.M.S.; Huda, D. A GIS Framework to Demarcate Suitable Lands for Combine Harvesters Using Satellite DEM and Physical Properties of Soil. *J. Geovisualization Spat. Anal.* **2023**, *7*, 27. [CrossRef]
51. Abrol, I.P.; Sangar, S. Sustaining Indian agriculture–conservation agriculture the way forward. *Curr. Sci.* **2006**, *91*, 1020–1025.
52. Shoba, P.; Ramakrishnan, S.S. Modeling the contributing factors of desertification and evaluating their relationships to the soil degradation process through geomatic techniques. *Solid Earth* **2016**, *7*, 341–354. [CrossRef]
53. Sahu, N.; Das, P.; Saini, A.; Varun, A.; Mallick, S.K.; Nayan, R.; Aggarwal, S.P.; Pani, B.; Kesharwani, R.; Kumar, A. Analysis of Tea Plantation Suitability Using Geostatistical and Machine Learning Techniques: A Case of Darjeeling Himalaya, India. *Sustainability* **2023**, *15*, 10101. [CrossRef]
54. Mondal, S.; Patel, P.P. Examining the utility of river restoration approaches for flood mitigation and channel stability enhancement: A recent review. *Environ. Earth Sci.* **2018**, *77*, 195. [CrossRef]
55. Ghimire, U.; Akhtar, T.; Shrestha, N.K.; Paul, P.K.; Schürz, C.; Srinivasan, R.; Daggupati, P. A Long-term Global Comparison of IMERG and CFSR with Surface Precipitation Stations. *Water Resour. Manag.* **2022**, *36*, 5695–5709. [CrossRef]
56. Ramya, M.; Sathiyamurthi, S. A raw data article on the physico-chemical properties of soil from six firkas in Dharmapuri district, Tamil Nadu, India. *Data Brief* **2022**, *43*, 108452. [CrossRef]
57. Jackson, M. *Soil Chemical Analysis*; Prentice Hall Inc.: Englewood Cliffs, NJ, USA, 1958; Volume 498, pp. 183–204.
58. Piper, C.S. *Soil and Plant Analysis*; Scientific Publishers: Elk Grove Village, IL, USA, 2019.
59. Walkley, A.; Black, I.A. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [CrossRef]
60. Olsen, S.R. *Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate*; US Department of Agriculture: Washington, DC, USA, 1954.
61. Stanford, G.; English, L. Use of the Flame Photometer in Rapid Soil Tests for K and Ca. *Agron. J.* **1949**, *41*, 446–447. [CrossRef]
62. Minasny, B.; McBratney, A. Digital soil mapping: A brief history and some lessons. *Geoderma* **2016**, *264*, 301–311. [CrossRef]
63. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef]
64. Marcot, B.G.; Hanea, A.M. What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput. Stat.* **2021**, *36*, 2009–2031. [CrossRef]
65. Yao, X.; Tham, L.G.; Dai, F.C. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [CrossRef]
66. Awad, M.; Khanna, R. Support Vector Regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers [Internet]*; Apress: Berkeley, CA, USA, 2015; pp. 67–80. [CrossRef]
67. Xu, Y.; Sun, J.; Zhang, J.; Xu, Y.; Zhang, M.; Liao, X. Combining AHP with GIS in synthetic evaluation of environmental suitability for living in China's 35 major cities. *Int. J. Geogr. Inf. Sci.* **2012**, *26*, 1603–1623. [CrossRef]
68. Chang, C.; Lin, C. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–39. [CrossRef]
69. Boehmke, B.; Greenwell, B.M. *Hands-on Machine Learning with R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.
70. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
71. Hu, L.-Y.; Huang, M.-W.; Ke, S.-W.; Tsai, C.-F. The distance function effect on k-nearest neighbor classification for medical datasets. *SpringerPlus* **2016**, *5*, 1304. [CrossRef]
72. Kuhn, M.; Johnson, K. Data pre-processing. In *Applied Predictive Modeling*; Springer Science Business Media: New York, NY, USA, 2013; pp. 27–59.
73. McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For. Ecol. Manag.* **2012**, *272*, 3–12. [CrossRef]
74. Pedregosa, F. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
75. Sharaff, A.; Gupta, H. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences: Proceedings of IC4S 2018*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 189–197.
76. Kumar, M.; Shaikh, V.R. Site Suitability Analysis for Urban Development Using GIS Based Multicriteria Evaluation Technique. *J. Indian Soc. Remote Sens.* **2013**, *41*, 417–424. [CrossRef]
77. Wang, H.; Zheng, H. True Positive Rate. In *Encyclopedia of Systems Biology*; Springer: New York, NY, USA, 2013; pp. 2302–2303. [CrossRef]
78. Bogunovic, I.; Mesic, M.; Zgorelec, Z.; Jurisic, A.; Bilandzija, D. Spatial variation of soil nutrients on sandy-loam soil. *Soil Tillage Res.* **2014**, *144*, 174–183. [CrossRef]
79. Wilding, L.P. Spatial Variability: Its Documentation, Accommodation and Implication to Soil Surveys. In Proceedings of the Soil Spatial Variability Proceedings of a Workshop of the ISSS and the SSA, Las Vegas, NV, USA, 30 November–1 December 1984; pp. 166–187.

80. Cooksey, R.W. Descriptive Statistics for Summarising Data. In *llustrating Statistical Procedures: Finding Meaning in Quantitative Data*; Cooksey, R.W., Ed.; Springer: Singapore, 2020; pp. 61–139. [CrossRef]
81. Soropa, G.; Mbisva, O.M.; Nyamangara, J.; Nyakatawa, E.Z.; Nyapwere, N.; Lark, R.M. Spatial variability and mapping of soil fertility status in a high-potential smallholder farming area under sub-humid conditions in Zimbabwe. *SN Appl. Sci.* **2021**, *3*, 396. [CrossRef]
82. Kılıc, O.M.; Ersayın, K.; Gunal, H.; Khalofah, A.; Alsubeie, M.S. Combination of fuzzy-AHP and GIS techniques in land suitability assessment for wheat (*Triticum aestivum*) cultivation. *Saudi J. Biol. Sci.* **2022**, *29*, 2634–2644. [CrossRef]
83. Montgomery, J.; Reid, M.D.; Drake, B. Protocols and structures for inference: A RESTful API for machine learning. In Proceedings of the Conference on Predictive APIs and Apps, PMLR, Boston, MA, USA, 10–12 October 2016; pp. 29–42.
84. Zhou, J.; Civco, D.L. Using genetic learning neural networks for spatial decision making in GIS. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1287–1295.
85. Dong, Y.; Zhang, Y.; Yue, J.; Hu, Z. Comparison of random forest, random ferns and support vector machine for eye state classification. *Multimed. Tools Appl.* **2016**, *75*, 11763–11783. [CrossRef]
86. Rajak, R.K.; Pawar, A.; Pendke, M.; Shinde, P.; Rathod, S.; Devare, A. Crop recommendation system to maximize crop yield using machine learning technique. *Int. Res. J. Eng. Technol.* **2017**, *4*, 950–953.
87. Xing, W.; Zhou, C.; Li, J.; Wang, W.; He, J.; Tu, Y.; Cao, X.; Zhang, Y. Suitability Evaluation of Tea Cultivation Using Machine Learning Technique at Town and Village Scales. *Agronomy* **2022**, *12*, 2010. [CrossRef]
88. Senagi, K.; Jouandeau, N.; Kamoni, P. Using parallel random forest classifier in predicting land suitability for crop production. *J. Agric. Inform.* **2017**, *8*, 23–32. [CrossRef]
89. Krawczyk, B. Learning from imbalanced data: Open challenges and future directions. *Prog. Artif. Intell.* **2016**, *5*, 221–232. [CrossRef]
90. Probst, P.; Wright, M.N.; Boulesteix, A. Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
91. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
92. Aguilar-Rivera, N.; Algara-Siller, M.; Olvera-Vargas, L.A.; Michel-Cuello, C. Land management in Mexican sugarcane crop fields. *Land Use Policy* **2018**, *78*, 763–780. [CrossRef]
93. de Carvalho, J.R.P.; da Silveira, P.M.; Vieira, S.R. Geoestatística na determinação da variabilidade espacial de características químicas do solo sob diferentes preparos. *Pesqui. Agropecu. Bras.* **2002**, *37*, 1151–1159. [CrossRef]
94. Alhamd, A.D.S.; Ibrahim, M.A. Unveiling soil and groundwater salinity dynamics and its impact on date palm yield in Southern Basrah, Iraq. *DYSONA-Appl. Sci.* **2024**, *5*, 25–32.


**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.