ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2024, 9 (2): 55-72 doi: 10.28978/nesciences.1569166

Machine Learning based Suggestion Method for Land Suitability Assessment and Production Sustainability

Yue Cao[1*], Dr. Liang Jiang[2*]

[1*] Krirk University, International College, Bangkok, Thailand.
E-mail: 9yuemoon@163.com

[2*] Krirk University, International College, Bangkok, Thailand.
E-mail: 736576870@qq.com

Abstract

The global population is projected to increase by an additional two billion by 2050, as per the assessment conducted by Food and Agriculture Management. However, the arable land is anticipated to expand by just 5%. Consequently, intelligent and effective agricultural practices are essential to enhancing farming production. Evaluating rural Land Suitability (LS) is a crucial instrument for agricultural growth. Numerous novel methods and concepts are being adopted in agriculture as alternatives for gathering and processing farm data. The swift advancement of wireless Sensor Networks (WSN) has prompted the creation of economical and compact sensor gadgets, with the Internet of Things (IoT) serving as a viable instrument for automation and decision-making in farmers. To evaluate agricultural LS, this study offers an expert system integrating networked sensors with Machine Learning (ML) technologies, including neural networks. The suggested approach would assist farmers in evaluating agricultural land for cultivating across four decision categories: very appropriate, suitable, somewhat suitable, and inappropriate. This evaluation is based on the data gathered from various sensor devices for system training. The findings achieved with the MLP with four concealed layers demonstrate efficacy for the multiclass categorization method compared to other current models. This trained system will assess future evaluations and categorize the land post-cultivation.

Keywords:

Agriculture, machine learning, geospatial analysis, three-dimensional space feature information integration, land suitability.

Corresponding Author: Yue Cao, E-mail: 9yuemoon@163.com

Article history:

Received: 09/05/2024, Revised: 06/07/2024, Accepted: 05/08/2024, Available online: 30/09/2024

Introduction

Agricultural farming is considered fundamental to human existence, serving as the principal source of sustenance and revenue for most nations globally (Mohamed et al., 2021). The nation's economy relies on agricultural output, which supplies food, raw resources, and jobs to its populace. Recent observations indicate a lack of substantial progress in crop yield within the farming sector (Agnolucci et al., 2020). There is a fast escalation in food prices due to crop output failing to satisfy demand. A contributing factor to the decline of agricultural production is farmers' reliance on conventional farming methods, resulting in diminished crop yields. The farmers who are inexperienced in agriculture possess an inadequate understanding of soil properties for growing crops (Selim, 2020). They are unaware that agricultural land requires evaluation before cultivation.

Land Suitability (LS) evaluation is an essential precondition for agricultural cultivation, facilitating optimal output (Nguyen et al., 2020). Farmers rely on soil testing laboratories to get soil parameters; however, these labs often need to be improved and deliver incorrect information. To get sufficient knowledge regarding farming, data must be gathered manually, posing significant challenges for farmers. The approach involves substituting conventional data-collecting techniques with sensors that utilize the Internet of Things (IoT) (Ali et al., 2020). Sensors are crucial for gathering data on numerous aspects, including soil, water, and climate, to enhance agricultural production. Utilizing data collected from various sensors, LS assessment is conducted, aiding farmers in assessing the present condition of their farmland and improving crop yields. Numerous decision-support systems have been created to help farmers make decisions on crop cultivation to optimize their profits.

Machine Learning (ML) algorithms assimilate knowledge from extensive datasets and seamlessly incorporate various data sources (Tripathi et al., 2024). Within the digital mapping system of land, these ML algorithms have been utilized to establish connections between soil measurements and additional factors, facilitating the comprehension of spatial and temporal variations in categories of soil and other soil attributes. This study acquired an agricultural dataset utilizing several IoT sensor gadgets, including a pH detector, soil moisture detector, salt sensor, and electromagnetic detector (Rajak et al., 2023). A detector is a device that identifies and reacts to specific inputs from the physical environment.

Due to its reduced labor requirements and time efficiency, it is employed in several real-time systems (Helo & Shamsuzzoha, 2020). Given the Internet's mediatory function in diverse interactions and information transmission, it is prudent to connect agricultural data with a cloud-based system. The data obtained from various IoT devices is saved on a cloud-based system.

The primary contribution of this paper is as follows:

- Identify pertinent features and acquire data from suitable sources for enhanced agricultural LS categorization, with the data being heterogeneous in structure.
- Consolidation of diverse data-collecting methods into a format conducive to future forecasting.
- The commendable effort to assist farmers in developing a viable approach, particularly in economically disadvantaged places, inside a country like Peru.
- The suggested model's generalization allows for application to any area since the existing test inputs adequately classify LS using this produced model.

Related Work

Incorporating the IoT into farming has resulted in the Internet of Agriculture of Things and sophisticated computer methodologies. The researchers implement this to maximize benefits and enhance agricultural productivity, Artificial Intelligence (AI) (Ben Ayed & Hanana, 2021), and the IoT. The agricultural sector is undergoing significant change and transformation driven by sensors, IoT, big data, and cloud technologies (Paraforos & Griepentrog, 2021). ML built an innovative farming system utilizing IoT coupled with a cloud surrounding, comprising four layers: data collecting, edge computing, transmission of data, and cloud-based computing. The IoT has been incorporated into agricultural systems to optimize profitability, with its applications classified into controlled environment management, open field cultivation, animal tracking, and aquaculture advancement (Mei et al., 2022).

A novel idea for agricultural applications was introduced by examining several integrated systems, including computing via the cloud, IoT, and data mining methodologies (Ali et al., 2023). A flexible network design was presented to monitor and govern agricultural fields in remote regions via IoT-based WiFi with long-range connectivity and fog computing capabilities. The evolution of IoT in farm data analysis has expanded from focusing on specific crops to encompassing all crop types. The framework could help with various applications, from managing and overseeing crops to marketing goods. Using the interpolation method, a fuzzy method was employed to construct an irrigation surveillance and management system that generates soil moisture level dispersion mapping (Baradaran & Tavazoei, 2022).

A prediction framework for precision agriculture was introduced utilizing IoT devices to enhance agricultural output cost-effectively (Ponnusamy & Natarajan, 2021). ML techniques were employed to predict illnesses in plants concealed inside leaf photos, achieving optimal classification accuracy.

A sophisticated method was created to forecast the fruit-melon picture of lesions on the skin and notify for novel planting situations utilizing convolutional neural networks and ML algorithms (Upadhyay & Kumar, 2022). The image is acquired using an infrared video sensor, achieving a commendable accuracy of 97.5%. The ZigBee Wireless Sensor Network (WSN) and ML techniques created an advanced autonomous agricultural surveillance system (Rahaman & Azharuddin, 2022). WSNs are used to track soil moisture and temperature levels. The method categorizes the moisture content of the soil.

A decision tree approach was presented to optimize both energy and water use. This was achieved with the aid of the IoT. Many soil factors were evaluated to assist farmers in conserving the diminishing water supply (Belachew et al., 2020). Soil moisture concentration and temperatures were employed to assess water requirements and inform choices on water delivery. The cloud-based IoT sensors facilitate the management of the entire system. This, therefore, conserves energy.

Three-Dimensional Space Feature Information Extraction

The evaluation comprised many steps, and a flowchart illustrating the processes employed in the study is presented in Figure 1. This study was performed in many phases:

- Identifying the sites for 100 soil profiles and obtaining soil samples.
- Evaluating the physicochemical characteristics of the soil samples.
- Gathering topographical and climatic characteristics.
- Assessing numerical ratings for soil, terrain, and climatic factors.
- Evaluate the LS indicator for grains.
- Determining the LS classification for grains.
- Establishing additional variables at consistent grid intervals.
- ML models established a correlation between supplementary factors and LS classification.
- Developing an ML-based LS classification map.
- Creating a conventional LS classification map.
- Contrasting the ML-based and traditional suitability of land classification maps.

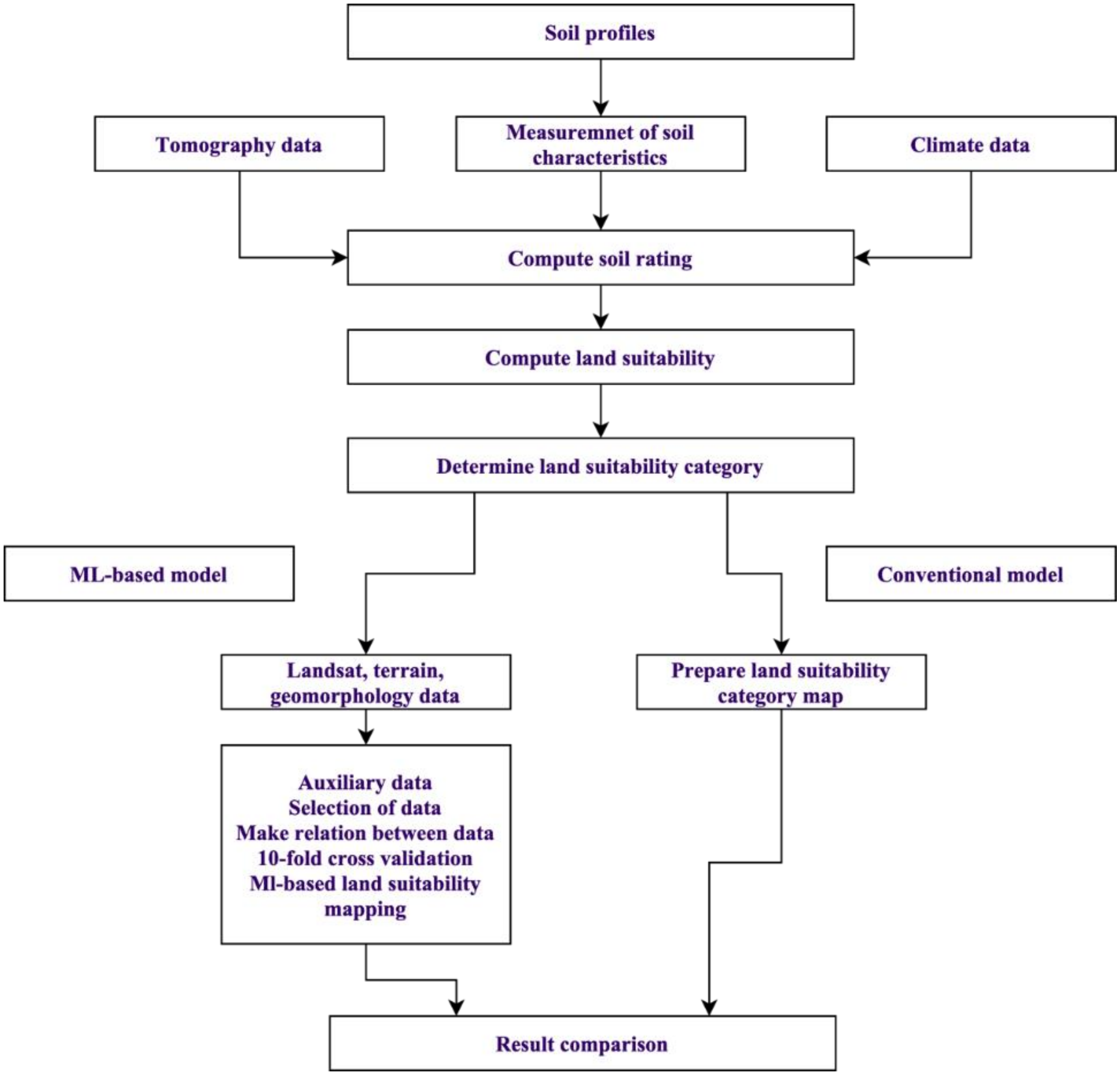


Figure 1. Workflow of the LS monitoring and management model

Geospatial features are the primary markers that distinguish Three-dimensional space feature information from other general information, and geometric relationships are meaningful quantitative representations of geospatial features, such as coordinates or location, proximity, shape, size, direction, and internal geometric structure. Compared with other relationships, the interrelationships of geospatial information entities are more complex and diverse, and the amount of data to describe them is much larger. Figure 2 is an example of a geospatial topological feature-containment relationship on the spatial topological feature relationships and intersection relationships; geospatial geometric features include coordinates, directions, angles, etc.

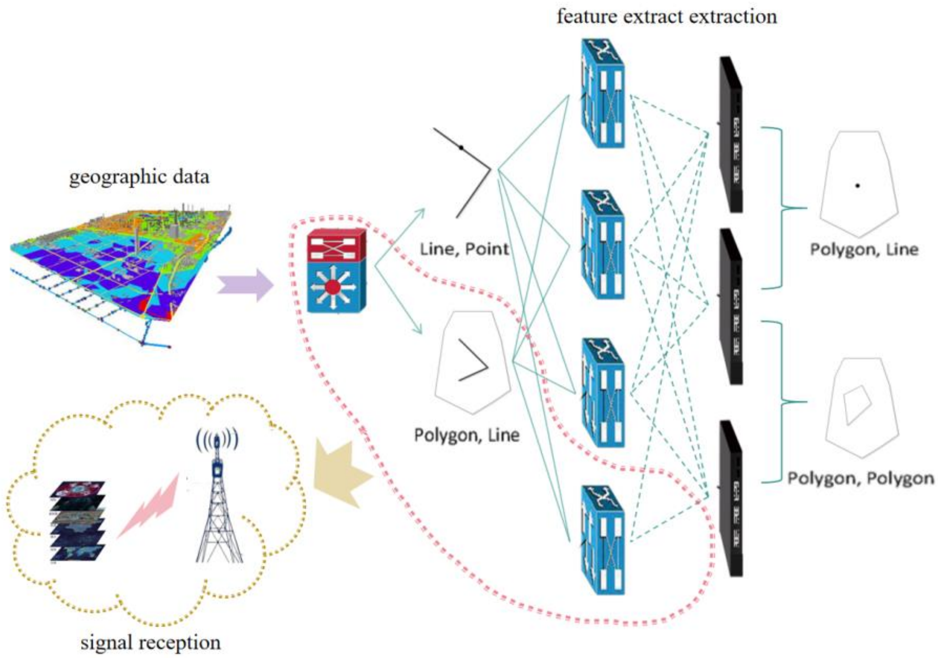
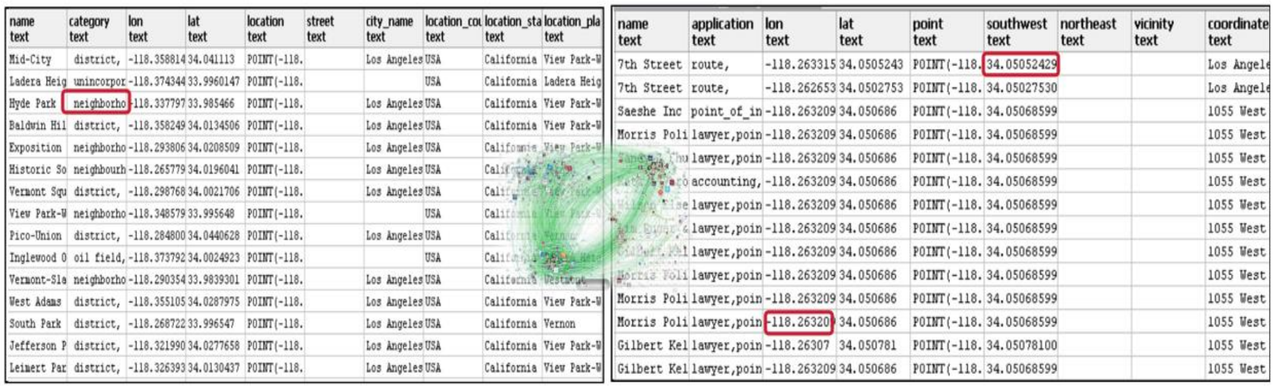


Figure 2. Geospatial topological outstanding feature

Extracting Three-dimensional space feature information from heterogeneous Three-dimensional space feature information data sources is a complex task. Most three-dimensional space feature information data sources provide visitors with appropriate access rights, allowing them to obtain original geographic data through the web interface provided by the data sources.

DBpedia extracts the URLs, longitudes, dimensions, and classes of all entities in a particular coordinate region in this way. The other data sources mentioned above are extracted through REST APIs. In this paper, the /place/nearby search method is used to get a list of locations within a range centered on a particular location, and the /place/details method can be used to get detailed information on a specific area. Due to the data source's limitation in data acquisition, it is impossible to acquire the whole dataset at once. Therefore, this paper takes a suitable approach to download the data in chunks and batches.

The structured markup language files of the raw data obtained from the data source are parsed to make the data more accessible to interpret and retain the integrity of the data. In data linking, it is necessary to use a cleaner and more effective data set. Therefore, it is essential to carry out data preprocessing, such as removing punctuation, garbled codes, suspended words, and other operations. Analyze the initial preprocessed data for Three-dimensional space feature information-related data attribute information such as building identification number, building name, classification, address, latitude, longitude, elevation, etc. Figure 3 shows some of the data selected after analysis.



name text	category text	lon text	lat text	location text	street text	city_name text	location_cou text	location_sta text	location_pla text
Mid-City	district,	-118.358814	34.041113	POINT(-118.		Los Angeles	USA	California	View Park-W
Ladera Heig	unincorpor	-118.374344	33.9960147	POINT(-118.			USA	California	Ladera Heig
Hyde Park	neighborho	-118.337797	33.985466	POINT(-118.		Los Angeles	USA	California	View Park-W
Baldwin Hil	district,	-118.358249	34.0134506	POINT(-118.		Los Angeles	USA	California	View Park-W
Exposition	neighborho	-118.293806	34.0208509	POINT(-118.		Los Angeles	USA	California	View Park-W
Historic So	neighbourh	-118.265779	34.0196041	POINT(-118.		Los Angeles	USA	[illegible]	[illegible]
Vermont Squ	district,	-118.298768	34.0021706	POINT(-118.		Los Angeles	USA	[illegible]	[illegible]
View Park-W	neighborho	-118.348579	33.995648	POINT(-118.			USA	[illegible]	[illegible]
Pico-Union	district,	-118.284800	34.0440628	POINT(-118.		Los Angeles	USA	[illegible]	[illegible]
Inglewood 0	oil field,	-118.373792	34.0024923	POINT(-118.			USA	[illegible]	[illegible]
Vermont-Sla	neighborho	-118.290354	33.9839301	POINT(-118.		Los Angeles	USA	California	Westmont
West Adams	district,	-118.355105	34.0287975	POINT(-118.		Los Angeles	USA	California	View Park-W
South Park	district,	-118.268722	33.996547	POINT(-118.		Los Angeles	USA	California	Vernon
Jefferson P	district,	-118.321990	34.0277658	POINT(-118.		Los Angeles	USA	California	View Park-W
Leimert Par	district,	-118.326393	34.0130437	POINT(-118.		Los Angeles	USA	California	View Park-W

name text	application text	lon text	lat text	point text	southwest text	northeast text	vicinity text	coordinate text
7th Street	route,	-118.263315	34.0505243	POINT(-118.	34.05052429			Los Angele
7th Street	route,	-118.262653	34.0502753	POINT(-118.	34.05027530			Los Angele
Saeshe Inc	point_of_in	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
Morris Poli	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
[illegible]	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
[illegible]	accounting,	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
[illegible]	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
[illegible]	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
[illegible]	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
Morris Poli	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
Morris Poli	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West
Morris Poli	lawyer,poin	-118.26320	34.050686	POINT(-118.	34.05068599			1055 West
Gilbert Kel	lawyer,poin	-118.26307	34.050781	POINT(-118.	34.05078100			1055 West
Gilbert Kel	lawyer,poin	-118.263209	34.050686	POINT(-118.	34.05068599			1055 West

Figure 3. Schematic representation of data analysis results

Three-Dimensional Space Feature Information Ontology Construction

One of the research goals of Three-dimensional space feature information ontologies is to form a common understanding of structured Three-dimensional space feature information between humans and software. Three-dimensional space feature information ontologies can separate generic domain knowledge from specific knowledge; ontological analysis and modeling of geographic domains can result in Three-dimensional space feature information ontologies, "a process that involves describing geographic domains in the objective world as a set of concepts and associations between them"; and Three-dimensional space feature information ontologies can help to analyze the Three-dimensional space feature information terminology's meanings.

A Three-dimensional space feature information ontology construction process consists of:1) determining the Three-dimensional space feature information ontology domain and scope;2) enumerating the essential concepts and terms in the Three-dimensional space feature information ontology;3) establishing the Three-dimensional space feature information ontology framework;4) defining classes and class hierarchies;5) defining attribute slots of classes and their fetch types and restrictions; and6) coding and formalizing the Three-dimensional space feature information ontology. Figure. 3 represents the Three-dimensional space feature information ontology constructed in this paper, showing the main components of the ontology (concepts, attributes, relationships, and hypotheses).

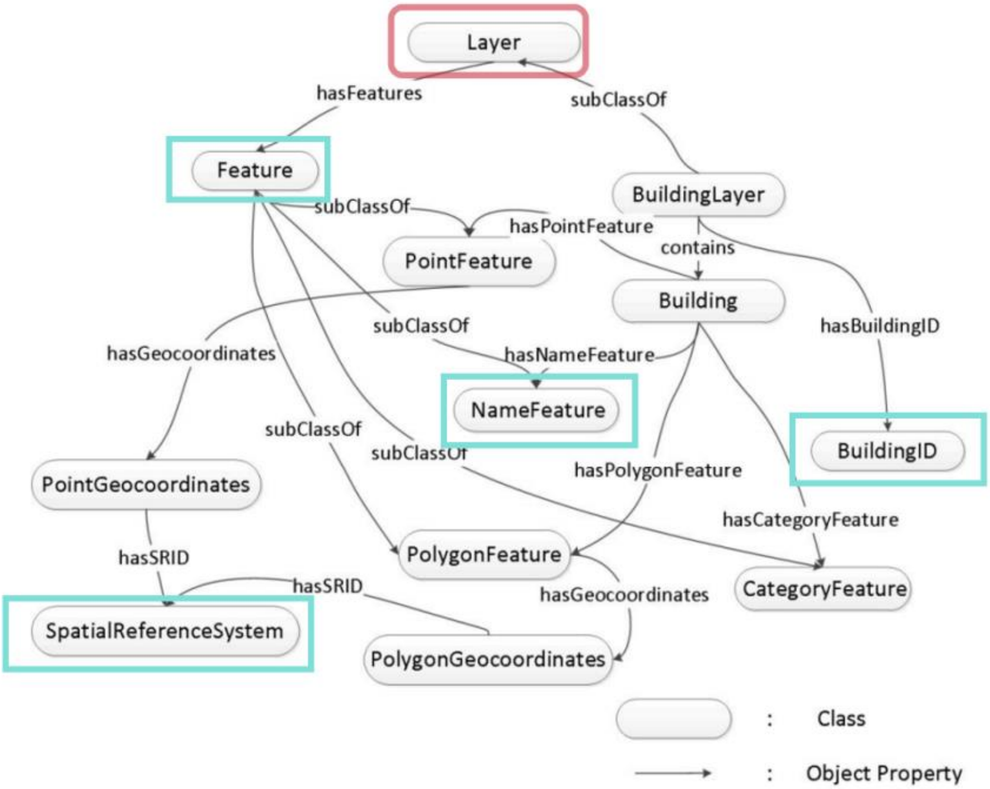


Figure 4. Three-dimensional space feature information ontology

This paper mainly focuses on buildings in the humanities and social entities as the research object and adopts three-dimensional space feature information on ontology construction. As shown in Figure 4, this is the ontology model constructed in the preliminary work, in which the geographic building entity is represented by the attributes and subclasses of the BuildingLayer class and the attributes and subclasses of the Feature class; in this paper, the SpatialReferenceSystem class is used to illustrate the identification of the geographic coordinate system of different information sources.

This paper uses the Karma tool to map the Three-dimensional space feature information ontology constructed above and the Three-dimensional space feature information data extracted from the information sources for information calibration. Mapping data to ontology using Karma is divided into two steps: describing the data types by assigning a syntactic type to each column and determining the relationship with the syntactic types inferred in the ontology. First, the appropriate syntactic types for the three-dimensional space feature information columns extracted from the heterogeneous data sources are defined and derived from the attributes of the three-dimensional space feature information ontology. Second, Karma automatically associates attributes in the ontology by specifying relationships between syntactic types. Users can easily adjust the correspondence between syntactic classes and modify the syntactic types through the graphical interface. Figure 5 shows the mapping relationship between OpenStreetMap data columns and Three-dimensional space feature information ontology.

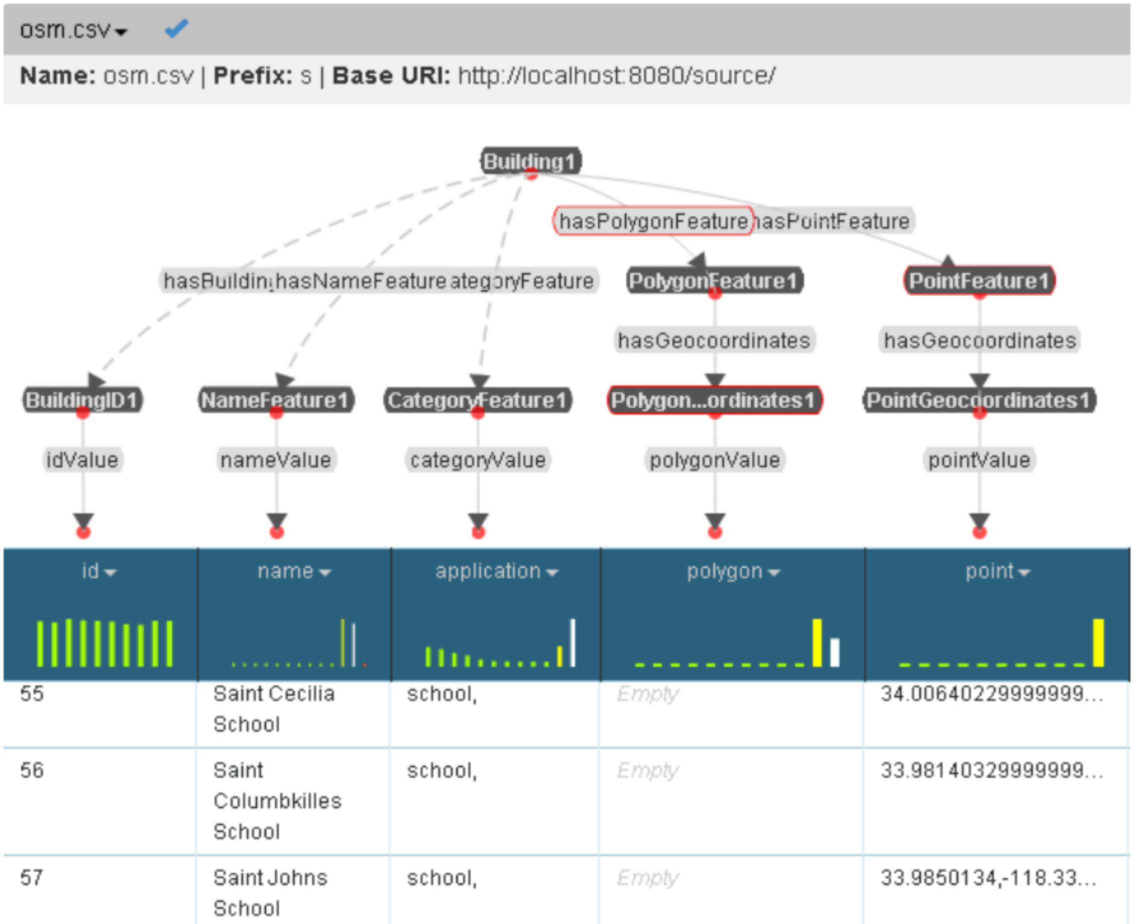


Figure 5. Mapping process using Karma

The black line in the middle is the object attributes of the Three-dimensional space feature information ontology, the gray line below is the data attributes of the Three-dimensional space feature information ontology, and the line below is the data labels of OpenStreetMap. We can see that Karma automatically maps the data labels in OpenStreetMap to the object properties in the Three-dimensional space feature information ontology, significantly improving the efficiency of mapping heterogeneous Three-dimensional space feature information. In addition, we can generate a model in RDF format from the mapping relationship and map the extracted geographic data into RDF data, and the saved mapping model can be used directly in future work.

Information Links based on Geospatial Features

Record linking refers to the identification of the same real things of different origins, which can solve the Different and residual quantities of data in different data sources, by which the quality of data and data integrity can be improved, and the cost and expense of data collection can be reduced. In previous methods for performing record linking, linking is accomplished by defining syntactic rules. However, by defining syntactic rules, geographic data's unique spatial relationship features are ignored, which play a vital role in linking. In addition, for the problem of different representation languages used by different information sources, complex preprocessing work has to be carried out before linking records. If the geospatial relationships of geographic data are utilized for record linking, the previous preprocessing work of translation in different languages can be avoided.

For the calculation of the maximum degree of correlation between existing geographic and physical names and the degree of similarity between categories of physical names, the Jaro-Winkler method and the Jaccard method were selected to calculate the degree of similarity between entity names and the degree of similarity between categories of entities, respectively. Since the known names themselves have a series of problems, such as simple methods of expression and expression of numbers, the method of calculating the degree of similarity of entity names needs to be able to accept more errors, and the Jaro-Winkler method, which is a method of calculating letters included based on the minimum writing distance, is in line with the requirements of the calculation of the degree of similarity. Based on the multi-feature similarity calculation results, the final extraction of the geospatial data is performed using the device autonomous learning KN classification method. The idea of the device autonomous learning KN classification method is as follows: first, the distance between the classification samples and the known category training samples is calculated by the degree of similarity, and the distance or degree of similarity is obtained with the closest K nearest neighbors between the samples to be categorized; then, based on the classes close to these classes to be categorized to determine the category of the classified data. Setting the K close regions of the samples to be categorized all belong to the same category, the samples to be categorized belong to the category; on the contrary, the scoring process is implemented for each candidate category, and the categories to which the samples to be categorized belong are judged based on the default operation law. $Ur(G_i)$ It is a binary operation with the following formula.

$$Ur(G_i) = \begin{cases} 1, if \ \ \mathrm{r}(G_i) \neq \varnothing \\ 0, otherwise \end{cases} i \in \{1,2,3\} \quad (1)$$

If record r has a Point attribute, then $Ur(G_1)$ =l; otherwise $Ur(G_1)$ =0. Similarly, if $Ur(G_2)$ =1 and $Ur(G_3)$ =1, record r has a Polyline attribute and a Polygon attribute, respectively.

Geospatial relationships in geographic data are essential features compared to other data. In the method proposed in this paper, three geospatial features supported by PostGIS, contain, overlap, and distance, are utilized for linking operations. Their expressions are as follows.

$$\text{isContained}\left(r_1\left(G_i\right), r_2\left(G_j\right)\right) = \begin{cases} \mathit{true}, \text{if and only if no points of } r_2\left(G_j\right) \\ \text{lie in the exterior of } r_1\left(G_i\right), \text{and at} \\ \text{least one point of the interior of } r_2\left(G_j\right) \\ \text{lies in the interior of } r_1\left(G_i\right) \\ \mathit{false}, \mathit{otherwise} \end{cases} \quad (2)$$

$$\text{is Overlap}\left(r_1\left(G_i\right), r_2\left(G_j\right)\right) = \begin{cases} \text{true, if } r_1\left(G_i\right) \text{ and } r_2\left(G_j\right) \text{ intersect,} \\ \quad \text{but one doesn't completely} \\ \quad \text{contain another} \\ \text{false, otherwise} \end{cases} \quad (3)$$

$$\mathit{distance}\left(r_1\left(G_i\right), r_2\left(G_j\right)\right) = (\text{float})\text{value}, i, j \in \{1,2,3\} \quad (4)$$

Formula 2 indicates whether there is a containment relationship between entities, which is not just a containment relationship of the former to the latter but can be a containment relationship of the latter to the former. Equation 3 indicates whether there is an overlapping relationship between the entities, and an overlapping relationship means that there is an intersection between the entities but not a containment relationship. Equation 4 indicates the distance between two geometric regions (centers), calculated based on the radius and radian of the earth's sphere.

Table 1 shows the pseudo-code of the geospatial feature-based Three-dimensional space feature information linking algorithm. S1 and S2 denote the data files with three-dimensional space feature information extracted from two data sources and mapped using the ontology described above.

Table 1. Information link codes

Input: RDF file: Sj, S_2.
Output: matchedPairList;
for all ry in S, and rz in S_2do

if(contained(r_1(G),r_2(G_1))=true or

contained(r_2(Gj),r_1(G_1))=true){

similarity(ri,r_2)←1.0.

matchedPairList.add(ri,r_2);

}else if(isOverlap(r_1(G),r_2(G))=true and

distance(r_1(G),r_2(Gj))<150){

$$\mathit{similarity}\left(r_1, r_2\right) = \left|1 - \frac{\mathit{distance}\left(r_1\left(G_i\right), r_2\left(G_j\right)\right)}{350}\right|$$

if(similarity(rj,r_2)>threshold){
matchedPairList.add(ri,r_2);

}

```
}else if(Ur,(G))=1 and Ur2(G1)=1 and
            distance(r(Gi),r2(G1))<150){
```

$$similarity\left(r_1, r_2\right) = \left|1 - \frac{distance\left(r_1\left(G_i\right), r_2\left(G_1\right)\right)}{350}\right|$$

```
  if(similarity(ri,r2)>threshold){matchedPairList.add(ri,r2);
  }
}
```

Firstly, the "Contain" relationship between records is calculated. If the compared entities fulfill the conditions, they are considered to represent the same entity, and the similarity between them is 1, which means that the link between these two entities is successful. This is because the geographic location of the same entity in different Three-dimensional space feature information sources should be the same or at least very similar. Second, if there is no "contain" relationship between the entities, the "overlap" relationship between the entities is compared. According to the previous experiments, it is found that if there is a cross-spatial relationship between the compared entities and the distance between the entities does not exceed 60 meters, then the "overlap" relationship between the entities will be the same. According to the previous experiments, if the spatial relationship between the compared entities is intersecting and the distance between the entities is not more than 60 meters, the degree of proximity between the entities is high.

On the contrary, their similarity is shallow when the distance between the entities exceeds 150 meters. Therefore, entity pairs with more than 150 meters between entities are directly discarded to improve computational efficiency. Equation 5 is to calculate the similarity between entities r and r. The constant 350 in the equation is obtained through experiments. Through experiments, it is found that among all the link matching pairs, when the distance between entities is no more than 60 meters, the accuracy of successful entity linking reaches 0.83, i.e., and the constant is approximately equal to 350. Of course, thresholds can be set freely according to the actual needs. There is a particular case when the compared entities only have a Point attribute; there is no "contain" and "overlap" relationship between them; the similarity between entities can be calculated by Equation 5.

$$similarity\left(r_1, r_2\right) = \left|1 - \frac{distance\left(r_1\left(G_i\right), r_2\left(G_j\right)\right)}{350}\right|, i, j \in \{1, 2, 3\} \quad (5)$$

Web-based Self-Directed Learning to Predict Consumption Distribution

This research utilized the conditioned Latin hypercube sample approach to allocate sites for 100 soil characteristics. The soil characteristics were characterized, specimens were obtained from genetic levels, air-dried at ambient temperature, and subsequently sieved using a 2 mm mesh before analyzing chemical and physical characteristics. The organic matter was quantified by wet burning.

Soil pH and electrical conductivity were measured in a saturating paste using a pH sensor and a conductivity gauge. The cation exchanging capability was assessed using the 1 N ammonium acetate technique at pH 7.0. The calcite carbonate equivalency was determined via a volumetric technique. The particle size dispersion was assessed using the Bouyoucos hydrometer technique. The convertible sodium proportion was calculated as the proportion of sodium to cation exchange capacity.

The gypsum level was determined; however, it wasn't considered zero. Topographical and climate information for LS evaluation were acquired from digitized elevation modeling and the Ghorveh synoptic meteorology stations for the 30 years spanning 1993 to 2023.

The suggested sensor-based AI algorithm assesses experimental agricultural land based on selected criteria for crop cultivation, which is anticipated to enhance production. The rural data collected from detectors comprises 1,000 options with 14 properties. Of the 1000 data situations, 750 are allocated for training models, while the other 250 are designated for evaluation of the method.

The testing dataset was gathered independently, following each crop yield for over one year. The suggested framework categorizes the data set under study into four judgment groups: most appropriate (category 1), pertinent (category 2), somewhat appropriate (category 3), and inappropriate (category 4). The agricultural site is classified as Category 1, while Category 2 is suitable for crop cultivation in its current state. Rural farms classified as Category 3 require treatment using suitable manure before growing crops, and land classified as Category 4 is unsuitable for farming. The assembled agricultural dataset is utilized to train the ML method. The efficacy of the ML-based multiple classes categorization model is juxtaposed with the outcomes derived from neuronal networks.

We perform a unified data extraction of Three-dimensional space feature information from multiple Three-dimensional space feature information sources and first calculate the similarity of geospatial relationships, names, and categories, respectively, to analyze the effect of single-feature matching on the results of Three-dimensional space feature information linking, as shown in Figure 6, in which the horizontal coordinates denote the IDs of entity pairs of data linking. The vertical coordinate is set to be the similarity value, in which the training results denote the similarity of 1 among the pairs of data-linking entities. Figure. 6(a) shows the effect of geospatial relationship similarity on data-linking results, Figure. 6(b) shows the effect of name similarity on data linking results and Figure. 6(c) shows the effect of category similarity on data linking results.

(a) Impact of interrelationships on data links

(b) Impact of name similarity on data linkage

(c) Impact of category similarity on data linkage

Figure 6. Effect of category Similarity on data linking results

To verify the accuracy of the multi-feature fusion method proposed in this paper for linking geographic 3D information data, the experimental results are evaluated from multiple perspectives by precision (P), recall by (R), and F-measure (β=0.5), as shown in Figure 7. Figure 7 compares this paper's method with the method proposed by Samal et al. Samal et al. proposed to utilize a graph theoretic approach to fuse geographic context-relevant and context-irrelevant information for multi-source feature matching. The method proposed by Samal et al. is implemented and applied to this paper's dataset. The results show that the method of this paper is significantly better than their method in terms of precision, recall, and F-value, which proves the feasibility and superiority of this paper's method; at the same time, the Support Vector Machines method and the K Nearest Neighbors method of classification algorithms are respectively combined with this paper's data linking method for comparison tests.

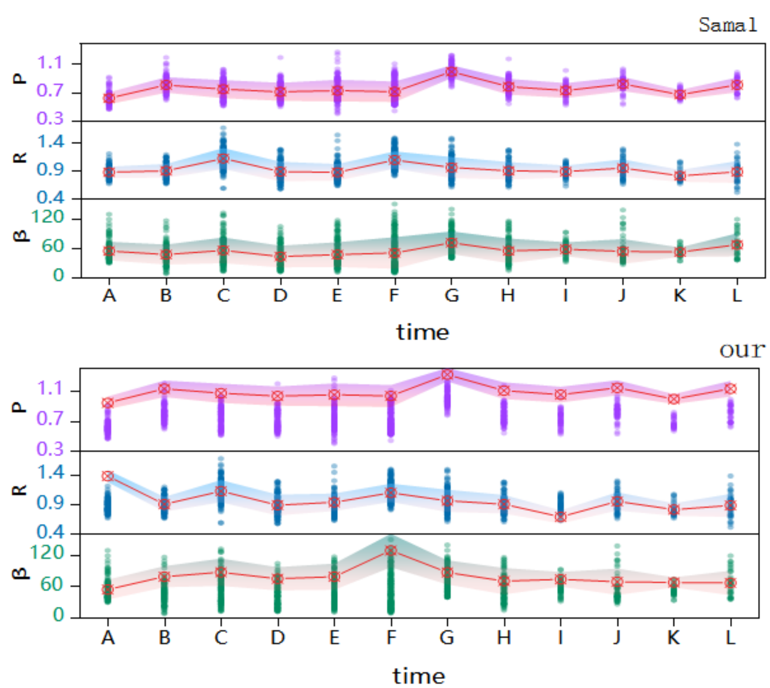


Figure 7. Model comparison validation

The geographic locations of 92 cities in 1985 were used as positive samples; 300 non-urban random points generated in the GIS environment were used for buffer analysis as negative samples, and the radius of the buffer zone of the 300 random points was set to 600m, making the total area of the positive and negative samples consistent. Using SPSS, 70% of the positive and negative samples were randomly selected as training set samples and the remaining 30% as validation set samples. Based on the results of correlation and logistic regression analyses, the selected influencing factors include elevation, slope, fault density, stream network density, buffer zones of deep fault zones, buffer zones of large-scale faults, buffer zones of streams of the first and second classes, and buffer zones of streams of the third class, buffers for first and second-order streams, and buffers for third-order streams. The number of decision trees (NT) was set to 300, and the Gini coefficient principle was used to generate decision trees for the stochastic fusion algorithm modeling. The validation set verified the generated model, representing the fusion decision model in Figure 8.

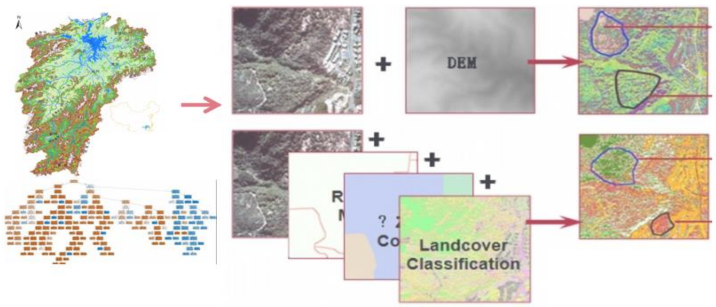


Figure 8. Fusion Decision Tree Model

Based on the stochastic fusion algorithm model, the study area is classified by the training set samples; in the study area range of consumption potential, as shown in Figure 9, the results are between 0 and 1, and the closer the value is to 1, it indicates that the potential of the area is better, the potential of goods consumption as shown in Figure 8. After the model construction is completed, the results are validated using the validation set samples, and the obtained estimate of the overall accuracy is 73.53%.

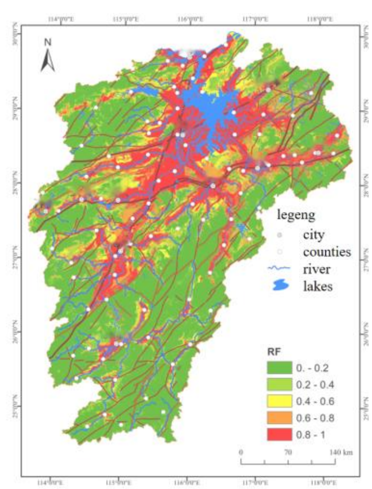


Figure 9. Consumption Potential of Goods in the Study Area

As the result of Figure. Eight shows: The plain around Poyang Lake, as well as the basins and valleys distributed along the Yangtze River, Gan River, Xin River, Yuan River, Jin River and Fu River basins, such as Jitai Basin, Ganzhou Basin, Fuzhou Basin and so on, are the most suitable areas to develop consumption; From the geological point of view, the fracture zones along the Ganjiang River, Pingxiang-Yi Guangfeng Fracture Belt, Yifeng-Jingdezhen Fracture Belt, Suichuan-Yi Linchuan Fracture Belt, Dayu-a Nancheng Fracture Zone, Yingtan-Anyuan Fracture Zone, and other super crystal fracture zones or deep fracture zones can be candidates for mainly stimulating the consumption of sporting goods.

To verify the correctness and practicability of the established comprehensive analysis model, the model and the traditional model were compared, respectively, from the iteration time and the number of iterations, as represented in Figure 10. The horizontal coordinate is the calculation time and the number of tests, respectively, and the vertical coordinate is the relative accuracy and the calculation accuracy, respectively; from the figure, it can be seen that, with the increase of iteration time, the calculation accuracy is maintained well and gradually increases, and with the rise of iteration times, the computational accuracy is significantly improved, which is 54.34% more accurate than the traditional model on average.

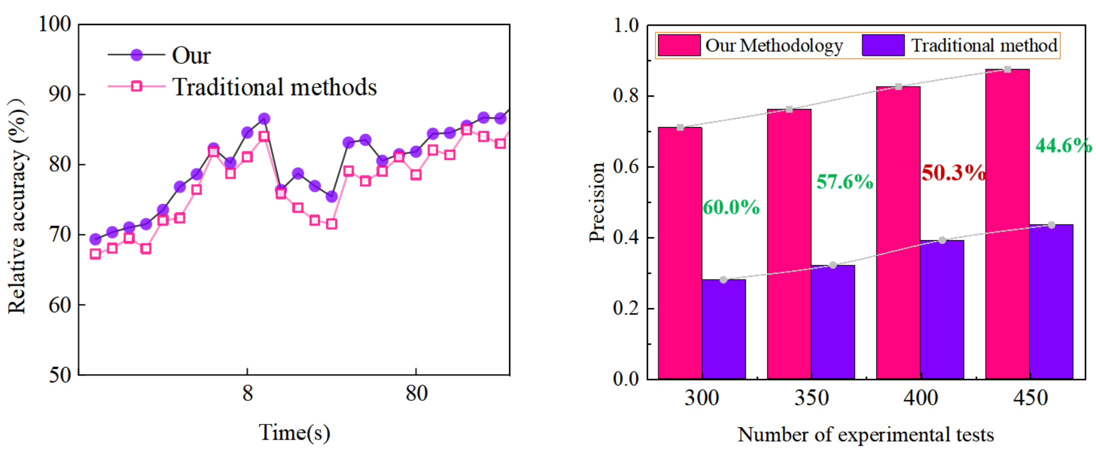


Figure 10. Comparison with the conventional model

Conclusion

Agriculture, the foundation of any nation, must ensure its sustained development. This study introduced a model with an accuracy of 99% that was deemed highly desirable. Information gathered from several sensors, processed using a multilayer perceptron with four concealed layers, has enhanced efficiency. A practical advice system with clear directives consistently yields superior outcomes. The reliability score relative to the accuracy score illustrates the efficacy of this suggested method, which guarantees proper categorization. Multiclass categorization in agriculture would enhance the suggestions system to provide precise guidance to farmers. Instead of binary categorization, this approach would give exact guidance to the producers. Consequently, this methodology would furnish real-time information to enhance agricultural output output.

Author Contributions

All Authors contributed equally.

Conflict of Interest

The authors declared that no conflict of interest.

References

Agnolucci, P., Rapti, C., Alexander, P., De Lipsis, V., Holland, R. A., Eigenbrod, F., & Ekins, P. (2020). Impacts of rising temperatures and farm management practices on global yields of 18 crops. *Nature Food*, *1*(9), 562-571.

Ali, A., Hussain, T., Tantashutikun, N., Hussain, N., & Cocetta, G. (2023). Application of smart techniques, internet of things and data mining for resource use efficient and sustainable crop production. *Agriculture*, *13*(2), 397. https://doi.org/10.3390/agriculture13020397.

Ali, I., Ahmedy, I., Gani, A., Munir, M. U., & Anisi, M. H. (2022). Data collection in studies on Internet of things (IoT), wireless sensor networks (WSNs), and sensor cloud (SC): Similarities and differences. *IEEE Access*, *10*, 33909-33931. http://dx.doi.org/ 10.1109/ACCESS.2022.3161929.

Baradaran, A. A., & Tavazoei, M. S. (2022). Fuzzy system design for automatic irrigation of agricultural fields. *Expert Systems with Applications*, *210*, 118602. https://doi.org/10.1016/j.eswa.2022.118602.

Belachew, A., Mekuria, W., & Nachimuthu, K. (2020). International Soil and Water Conservation Research. *International Soil and Water Conservation Research (ISWCR)*, *80*. https://doi.org/10.1016/j.iswcr.2020.01.005.

Ben Ayed, R., & Hanana, M. (2021). Artificial intelligence to improve the food and agriculture sector. *Journal of Food Quality*, *2021*(1), 5584754. https://doi.org/10.1155/2021/5584754.

Helo, P., & Shamsuzzoha, A. H. M. (2020). Real-time supply chain—A blockchain architecture for project deliveries. *Robotics and Computer-Integrated Manufacturing*, *63*, 101909. https://doi.org/10.1016/j.rcim.2019.101909.

Mei, Y., Sun, B., Li, D., Yu, H., Qin, H., Liu, H., & Chen, Y. (2022). Recent advances of target tracking applications in aquaculture with emphasis on fish. *Computers and Electronics in Agriculture*, *201*, 107335. https://doi.org/10.1016/j.compag.2022.107335.

Mohamed, E. S., Belal, A. A., Abd-Elmabod, S. K., El-Shirbeny, M. A., Gad, A., & Zahran, M. B. (2021). Smart farming for improving agricultural management. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 971-981.

Nguyen, H., Nguyen, T., Hoang, N., Bui, D., Vu, H., & Van, T. (2020). The application of LSE software: A new approach for land suitability evaluation in agriculture. *Computers and Electronics in Agriculture*, *173*, 105440. https://doi.org/10.1016/j.compag.2020.105440.

Paraforos, D. S., & Griepentrog, H. W. (2021). Digital farming and field robotics: Internet of things, cloud computing, and big data. *Fundamentals of Agricultural and Field Robotics*, 365-385. https://doi.org/10.1007/978-3-030-70400-1_14.

Ponnusamy, V., & Natarajan, S. (2021). Precision agriculture using advanced technology of IoT, unmanned aerial vehicle, augmented reality, and machine learning. *Smart Sensors for Industrial Internet of Things: Challenges, Solutions and Applications*, 207-229. https://doi.org/10.1007/978-3-030-52624-5_14.

Rahaman, M. M., & Azharuddin, M. (2022). Wireless sensor networks in agriculture through machine learning: A survey. *Computers and Electronics in Agriculture*, *197*, 106928. https://doi.org/10.1016/j.compag.2022.106928.

Rajak, P., Ganguly, A., Adhikary, S., & Bhattacharya, S. (2023). Internet of Things and smart sensors in agriculture: Scopes and challenges. *Journal of Agriculture and Food Research*, *14*, 100776. https://doi.org/10.1016/j.jafr.2023.100776.

Selim, M. M. (2020). Introduction to the integrated nutrient management strategies and their contribution to yield and soil properties. *International Journal of Agronomy*, *2020*(1), 2821678. https://doi.org/10.1155/2020/2821678.

Tripathi, A., Waqas, A., Venkatesan, K., Yilmaz, Y., & Rasool, G. (2024). Building flexible, scalable, and machine learning-ready multimodal oncology datasets. *Sensors*, *24*(5), 1634. https://doi.org/10.3390/s24051634.

Upadhyay, S. K., & Kumar, A. (2022). A novel approach for rice plant diseases classification with deep convolutional neural network. *International Journal of Information Technology*, *14*(1), 185-199. https://doi.org/10.1007/s41870-021-00817-5.