International Journal of Engineering Research and Development

DOI: 10.29137/ijerad.1468766

Volume:17 Issue:01 Pages:92-104 March/2025

Research Article

Wireless Communications Standard Recommendation System Design for Smart Agriculture Applications

Ahmet Yazar[1*], **Ayşe Rabia Soylu**[2], **Orçun Balatlıoğlu**[2]

[1]*Eskişehir Osmangazi University, Department of Software Engineering, 26040 Eskişehir, TÜRKİYE*
[2]*Eskişehir Osmangazi University, Department of Computer Engineering, 26040 Eskişehir, TÜRKİYE*

***Received**: 15/04/2024* ***Accepted**: 09/10/2024* ***Published Online:** 15/03/2025*
***Final Version**: 01/03/2025*

Abstract
After the widespread adoption of industrial agricultural systems, the development of smart farming systems has gained momentum, especially in the last decade. Within the scope of this acceleration, wireless communication systems have assumed a pivotal role in the field of smart agriculture. The realization of the Internet of Things (IoT) concept in the context of smart farming has become possible in conjunction with next-generation wireless communication systems. Particularly, smart farming applications emerged as a significant use case during the standardization of 5th Generation (5G) cellular communication systems. Apart from 5G, there are different wireless transmission standards with advantages that can be leveraged in smart farming systems. In this study, a recommendation decision system has been developed to determine which wireless communication standard would provide higher benefits under various dynamic scenarios, taking into account the dynamic nature of different subsystems desired for smart farming. This approach, developed using machine learning, enables the decision-making process regarding which wireless communication standards can be utilized more effectively in smart farming applications. The benefit of working in real-world scenarios is demonstrated by including an example case study within the scope of the paper.

Keywords
5G, 6G, Smart agriculture, wireless communications, machine learning, recommendation system.

**** Corresponding Author:** ahmtyzr@gmail.com*

1. Introduction

In industry, agriculture and logistics systems, where speed is one of the key criteria, new business ideas planned to meet constantly changing consumer demands and needs need to be implemented faster. However, consultancy services can be expensive during the planning phase of the business idea. Choosing the wrong technologies during the research of available technologies and determining the most suitable one causes time and financial losses. This study was conducted for people who want to use their resources for smart agriculture, but may be not capable in analyzing which smart agriculture systems to install at what scale and how to provide communication between these systems. The developed approach was studied with this motivation.

There are various standards that stand out with their different features in wireless communication systems (Hossain and Markendahl, 2021). The areas of use of communication systems designed based on these standards may vary. Different wireless communication systems may be needed in smart agriculture systems (Ayaz et al., 2019). However, considering the need for systems that automatically produce suggestions for which communication standard can be used in which situations, the approach presented in this study was developed. In this context, a recommendation system was designed by using machine learning algorithms. The problem definition considered is presented in summary in Figure 1.

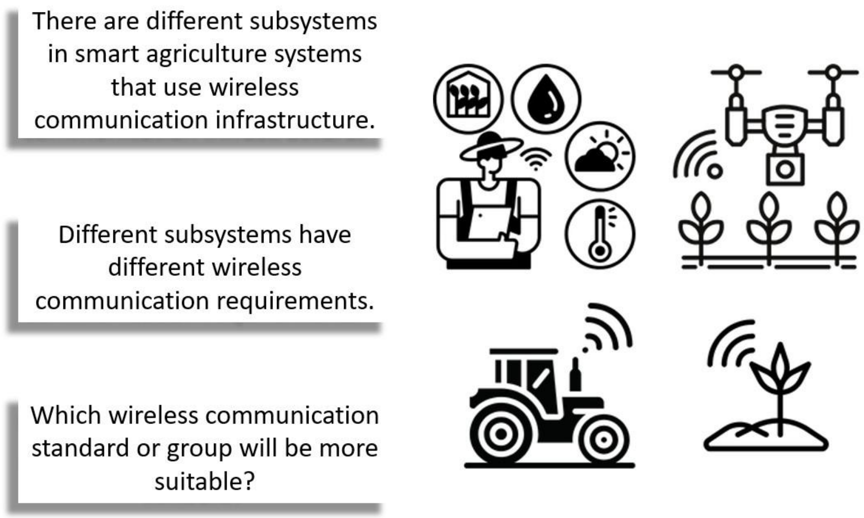


Figure 1. A graphical summary for the problem definition

In the development process of 5th Generation (5G) cellular communication systems, smart agriculture scenarios have also played a role, and the communication requirements of the sub-components in smart agricultural systems have been taken into account. It is anticipated that this role will also affect the development process of 6th Generation (6G) communication systems (Author et al., 2020; Author, 2021). When we look at the current literature, it is seen that a method that suggests a communication standard according to the changing needs in smart agricultural systems in this study has not been developed. However, in the study (Khan et al., 2023), different problems and solution suggestions regarding how 5G communication systems can be used in smart agriculture have been examined. What 5G and digital twin technologies can bring in terms of smart agricultural systems has been discussed in a different study (Fuentealba et al., 2022). How 5G and Internet of Things (IoT) structures can be used in smart agricultural systems in terms of sustainable cities has been the subject of one study (Jape et al., 2023). Another study examined what is expected in smart and sustainable cities where 6G can be used, including smart agriculture scenarios (Ouafiq et al., 2022). The other study addresses the convenience provided by the combined use of 5G and IoT devices in milk production and animal husbandry (Zhang et al., 2021). Another study examined the general characteristics of low-power wide area network (LPWAN) technologies and explained why they are important for smart agriculture (Liya and Arjun, 2020). In a recent study, automation-based technologies for irrigation systems in smart agriculture were examined (Arık and Korkut, 2022).

In this study, smart agriculture systems were examined, and by including different scenarios and the geographical conditions of the relevant region, a machine learning-based method was designed that suggests a wireless communication standard suitable for these scenarios and region conditions. In addition, the products and services of companies in the smart agriculture sector were investigated and a comprehensive scenario catalog was created. In addition, since there is no open-source dataset on the subject in the literature and various data providers, a new dataset was created by performing a requirement analysis according to the communication standards of smart agriculture scenarios.

Although there are no studies in the literature on the problem of determining wireless communication standards for smart agriculture applications, the synthetic dataset generation method based on the relationships on the wireless communication channel features used in this study has been previously examined in different studies (Hançer and Yazar, 2023a; Hançer and Yazar, 2023b; Eren et al., 2023; Sazak and Yazar, 2023; Altunan et al., 2023). A dataset for determining the appropriate waveform (Hançer and Yazar, 2023a; Hançer and Yazar, 2023b), a dataset for determining the appropriate vehicle-to-everything communication protocol (Eren et al., 2023), and a dataset for numerology decisions (Sazak and Yazar, 2023; Altunan et al., 2023) are the approaches developed in the studies in the literature. Due to the different problem definitions and application scenarios in the existing studies mentioned, the features and class labels have changed. Therefore, an artificial dataset was created using different features and class labels in this study. However, similar structures in the example studies in the literature were preferred for the dataset production model.

In Section 2, the details of the developed design are given under the title of material and method. In this context, smart agriculture scenarios and the created dataset are also processed in the same section. In Section 3, the findings and discussions regarding the machine learning results are presented. Finally, conclusions are given and future plans are exemplified in Section 4.

2. Material and Method

Under this section, firstly smart agriculture scenarios are discussed. What kind of wireless communication needs may arise under different smart agriculture scenarios are examined. Then, how the dataset is created is explained, and finally, details of the developed approach and method are presented.

2.1. Smart Agriculture Scenarios

In the studies conducted to obtain information about smart agriculture scenarios, our aim is to determine which machines and sensors are used for what purposes in smart agriculture. The first subject to be discussed is the scenarios in which smart agriculture is used. As seen in Table 1, three main smart agriculture scenario categories have been determined. The first of these scenarios is monitoring. In the scenarios where monitoring is performed, the main purpose is to make observations in the area where agriculture is performed. Thanks to these observations, the needs related to soil and plants are determined by the person doing the agriculture. The aim of the control scenarios, which are the second smart agriculture scenario category, is to provide the person doing the agriculture with the opportunity to remotely examine and control that area through machines. In the scenarios in the third and last scenario category, the tracking category, data is received from the target to be followed at short time intervals. Thanks to this data, the user can view the exact location of the source he wants to follow in real time.

Table 1. Smart Agriculture Scenarios and Groupings

Monitoring	Control	Tracking
Environment	Irrigation	Animal Movements
Humidity	Harvesting and Collection	Land Location
Temperature	Fertilization	
Growth	Pest	
Quality	Weed	
Disease	Treatment Plans	
Phenotype	Problem Identification	
Plant Count	Sorting and Packing	
Live Stock	Thinning and Pruning	
Animal Health	Sheep Herding	
	Milking	
	Offspring Care	

The scenarios and methods of choosing robotic devices used within different agricultural scenarios may vary. When we look at the application examples, it is not always possible to use robots in collecting and harvesting for the harvest and collection scenario. Sensitive crops such as soft fruits should be collected by hand. Other products can be collected using robotic devices. For the weed control scenario, there is a method called micro spray, which uses chemicals together with robots to kill weeds. In this method, a land vehicle with a camera on the field and image processing technology are used. Weeds detected in the processed image can be sprayed with pinpoint chemical spray. Other methods also have camera and image processing technology. However, physical methods such as laser are used instead of chemicals. In the weed scenario, the areas where weeds are located can be detected with the help of image processing technology in images taken with drones. There is a method called active form recognition method. Thanks to this method, machines can identify up to 19 weed species. In the sorting and packaging scenario, separation and packaging is a delicate task. Thanks to robots, this task can be performed continuously and uninterruptedly. Robots with human-like eye-hand coordination skills separate and pack loose parts from a moving conveyor using an image processing system. For the fertilization and irrigation scenario, ground robots that can navigate above the soil can autonomously navigate between crop rows and pour water directly to each base of the plant. Robots also have the advantage of accessing hard-to-reach areas. In the thinning and pruning scenario, parts that need to be pruned are cut by passing over the plant with the help of image processing technology. For the phenotype analysis scenario, the data obtained with drones and robots provides ease of setting the conditions that will achieve the highest yield by creating a genetic map of the crop. In the water-mineral monitoring scenario, LiDAR-supported robots can perform instant mineral and water amount analysis. For the treatment plans scenario, conditions such as whether the crops monitored with the help of drones have diseases and are under stress can be diagnosed early. After this detection, a treatment plan can be created for the crop. In the milking scenario, robots are used to help provide hygienic environments during the process. In addition to performing the milking process itself, it prepares the cows for the milking process and constantly controls the amount of milk produced by the cow. It helps to keep this amount at optimal values. Finally, in the sheep herding scenario, drones constantly monitor the herd from the air and prevent any sheep from leaving the herd.

When these scenarios and application examples are examined in general, the use of robot and drone systems in agriculture and animal husbandry is widespread. The reason for this is that it provides great convenience to the farmer in many areas. However, it is not possible to use these devices alone. Therefore, a network structure must be established in the agricultural area. The main purpose of this study is to determine how this network structure can be created in the most efficient way. In addition, bio sensors, pH sensors, passive infrared sensors, humidity sensors, pressure sensors, irrigation sensors, temperature sensors and color sensors can be used in smart agricultural scenarios. As can be seen, there is a strong unity between sensors and agricultural scenarios. In addition, sensors need to collect the data they obtain at a central point. At this point, there is more than one communication standard technology to ensure data transmission.

2.2. Forming The Dataset

There is no existing dataset in the literature regarding the examination of wireless communication standards within the scope of the research topic in this study and smart agriculture scenarios. At this point, it was desired to produce a suitable dataset, and due to the difficulties of creating a measurement-based dataset, a synthetic dataset was preferred (Emam, 2020). In recent years, synthetic dataset production can be preferred for very different applications in the literature (Nikolenko, 2022). There are different advantages of using synthetic datasets. For example, in order to perform machine learning against very rare situations in reality, the desired number of these rare situations can be produced. A negative aspect of artificial datasets is that simulation studies must be done in detail to create a realistic dataset.

During the creation of the artificial dataset, situations where the wireless communication channel features are heavily affected were taken into account. The features to be used were determined within this scope. In particular, the change in environmental conditions significantly changes the communication channel effects (Kihero et al., 2021; Yarkan and Arslan, 2008). The path loss effect changes under different environmental conditions (Kihero et al., 2021); multipath effects also vary depending on environmental factors (Yarkan and Arslan, 2008).

In the process of creating the dataset, the flow in the block diagram shown in Figure 2 was carried out on a computer simulation. For the production of the synthetic dataset, first of all, class information was created (Emam, 2020), and features were created by adding different randomness within certain limits over the class information. The parameters for the features given in Table 2 for different class labels were produced by considering the Normal distribution between predetermined limit values. At this point, the relationships and dependencies between different parameters were also taken into account. In the next stage, feature values were obtained from the input parameters produced in a way that would be normalized between 1-10 values. In the last stage, the features were recorded together with the class information and a sample dataset was created for the target.

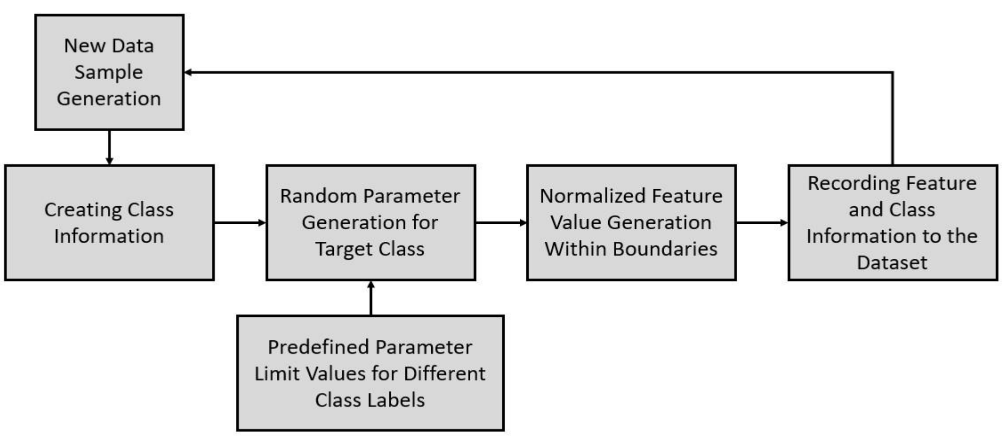


Figure 2. Block diagram representation of the method used for dataset generation

Different smart agriculture applications; consumer, farmer, operator and investor demands; changing land and environment characteristics create direct communication system requirements. The created model tries to determine the most suitable one among the wireless communication standard groups wide area networks (WAN), local area networks (LAN) and low-power wide area networks (LPWAN) in line with the system inputs. These standard groups also created class labels for the creation of classification models.

The system inputs that change depending on the different smart agriculture scenarios and land characteristics explained in the previous section are defined as the features of the dataset that the machine learning algorithms will learn from, as can be seen in Table 2. Detailed explanations of these features are given below. Wireless communication channel relationships are also presented in this context. During data production, production was carried out by taking wireless communication channel relationships into account.

Table 2. Features as the System Inputs

No	Feature Variables in the Algorithm	Scaled System Input	Equivalent of Feature Lower Limit in System (1)	Equivalent of Feature Upper Limit in System (10)
1	Region Geography	Physical characteristics of the land	A simple geography (plain, etc.)	A difficult geography (hillside, forest, etc.)
2	Region Type	Rural/urban measure of the land	A rural area (low floors, etc.)	An urban area (high floors, etc.)
3	Land Size	Size of the land	Very small	Very large
4	Vehicle Density	Amount of actively moving vehicles and machines	Low density	Very dense
5	Average Mobility of Vehicles	Mobility measure of actively moving vehicles and machines	Very low	Very high
6	Number of Camera Systems	Number of devices transmitting images	A few	A large number
7	Data Throughput	Volumetric measure of transmitted data	Low criticality	High criticality
8	Reliability	Compensability of data loss	Low criticality	High criticality
9	Latency	Compensability of data delay	Low criticality	High criticality
10	Sensor Density	Amount of sensors sending small-sized data	Low density	Very dense
11	Sensor Data Transmission Frequency	Data transmission frequency of sensors	Scale	Continuous

Feature-1 It is related to the Regional Geography feature on Table 2. It was used to classify the region to be used as a smart agricultural land in terms of physical features such as whether it is mountainous, plain, forested and distance to water resources. This parameter was required because inputs such as height of buildings and density of trees create negative effects on the wireless communication channel due to multipath (signal reflections). As the physical features of the terrain become more difficult, multipath effects increase in terms of the wireless communication channel. For example, in a mountainous terrain, due to the high multipath effects, the amount of interference between symbols will be high and communication reliability will decrease. In the opposite case, for example, in a plain and desert-like terrain structure, only a single-reflection channel model will be encountered in terms of multipath effects, therefore wireless communication reliability will be expected to be higher. The importance of this feature changes according to the level of uninterrupted communication need from wireless communication requirements. The data model was designed to work only on agricultural land. In line with this goal, the geographical difficulty level is shown with Feature-1 added to the dataset. Among the criteria used, the number of trees per square kilometre in the land, the number of hills in the land and the slope of the location of the land were taken into account.

Feature-2 It is related to the Region Type feature on Table 2. It is used to classify agricultural land in terms of rurality and urbanity. This feature, which will affect the selection of the wireless communication channel model that can be used, also gives a clue about the distance to the nearest base station, i.e. the sender, within the land. In more urban areas, communication channel effects are more intense and problematic. In addition, in rural areas, communication infrastructure is generally planned to be placed more sparsely. Therefore, the importance of this feature changes in relation to both the need for uninterrupted communication and the distance criterion to the communication infrastructure.

Feature-3 It is related to the Land Size feature in Table 2. It is used to scale the size of the smart agricultural land. It is assumed that the number of vehicles and sensors transmitting data increases as the size of the land increases. This increase highlights the need for high connectivity in terms of wireless communication. In addition, path loss effects are greater for large lands. The default values of the land size feature while creating the dataset are exemplified in Table 3. While classifying, the maximum coverage area of the Sigfox technology, which is the LPWAN technology, was determined as the upper limit. The reason for using this technology as the upper limit is that it has the widest coverage area among LPWAN technologies. As the land size decreases, the need to choose between communication standards decreases, and vice versa.

Table 3. Range Values Used for Terrain Size Feature

Value of Feature in Dataset	Land Size (km^2)
1	0 – 3.5
2	3.6 - 7
3	7.1 – 10.5
4	10.6 - 14
5	14.1 – 17.5
6	17.6 – 21
7	21.1 – 24.5
8	24.6 - 28
9	28.1 – 31.5
10	31.6 - 35

Feature-4 It is related to the Vehicle Density feature on Table 2. The number of vehicles that are actively working on the field, such as irrigation and fertilization vehicles, is scaled. These vehicles will require more bandwidth compared to sensors and they also use communication links more intensively. While there is a single vehicle system per km2 in the field in the low-density case, it is assumed that there are 10-15 vehicles per km2 in the very dense case. When the vehicle density increases, the need for a WAN or LAN standard group may be greater. In addition, the use of information on how many mobile systems are is considered important in terms of Doppler effects (frequency shifts depending on speed) in the communication channel. In other words, this feature also affects Feature-5.

Feature-5 It is related to the Average Mobility of Vehicles feature on Table 2. It is considered as a measure of how fast the positions of the vehicles used on the field change. When it is assumed that the vehicles on the field will be mobile, the wireless communication channel is exposed to different negativities due to Doppler effects. At this point, average mobility information will be used. If there is vehicle mobility, Doppler effects will be higher, and when combined with multipath effects, the amount of interference between carriers will increase. In this case, if smart agriculture application needs higher uninterrupted communication will be required. Therefore, a communication standard may be needed among WAN or LAN standards.

Feature-6 It is related to the Number of Camera Systems feature on Table 2. It was used to scale the general need for image transmission in smart agriculture systems in different scenarios. Systems that include remotely controlled vehicles such as drones require high data volume, fast data path and reliable (robust) communication link. If the number of camera systems is high for the relevant smart agriculture scenario, LPWAN features are very insufficient at this point. Therefore, the WAN group comes to the fore. Both high data throughput, high data speeds and reliable communication needs can be provided with WAN group standards. If the number of camera systems is low, it is necessary to focus on other requirements and there is no superiority.

Feature-7 It is related to the Data Throughput feature in Table 2. It is used to scale how often the sensors in the system installed in the field should send data to a remote center. If there is a need for very frequent data transmission, in other words, if there is a need for continuous data transmission, other alternatives come to the fore rather than LPWAN standards. On the other hand, for example, in the case of a data transmission frequency of once per second or less, LPWAN standards come to the fore. The values taken here are scaled between once per day for low data frequency and once per millisecond for high data frequency.

Feature-8 It is related to the Reliability feature in Table 2. It is a measure of how much the loss of information transmitted by various sensors and devices can be compensated. If the reliability requirement is critical, a standard group that can respond better to the adverse conditions of the wireless communication channel should be preferred. Otherwise, a prioritization should be made based on other features and requirements to make a choice among the standards.

Feature-9 It is related to the Latency feature in Table 2. It is a measure of the importance of real-time transmission of information transmitted by various sensors and vehicles. In scenarios involving fault detection and drones, communication delays are expected to be low. Therefore, focus should be on communication standards that can provide lower latencies. If latencies are not critical, other features and requirements should be taken into account for prioritization in the selection of standards.

Feature-10 It is related to the Sensor Density feature in Table 2. It is a measure of the amount of sensors transmitting data, which varies depending on the size of the agricultural land. It is used to determine how much IoT communication systems are needed. The higher the sensor density, the more IoT communication systems there will be, and therefore LPWAN standards will come to the fore. On the other hand, if the sensor density is low, it is likely that other standard groups will come to the fore.

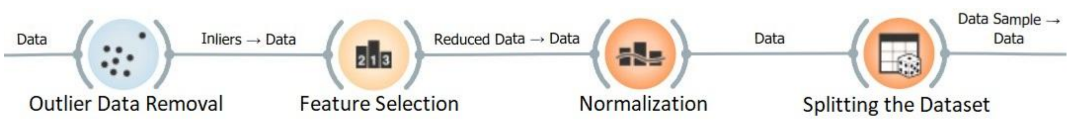
Feature-11 Table 2 is related to the Sensor Data Transmission Frequency feature. It is used to scale the volume of data that the sensors used in the smart agriculture scenario are expected to transmit to a remote server. As continuity increases, communication standards for higher data volumes come to the fore. On the other hand, if there is a more intermittent communication, the data volume from the sensors will be evaluated as quite low, and the most basic LPWAN solutions will become prominent.

Among these given features, three of them (Area Geography, Area Type and Land Size) vary depending on the user's land. The remaining eight features (Vehicle Density, Average Mobility of Vehicles, Number of Camera Systems, Data Throughput, Reliability, Latency, Sensor Density and Sensor Data Transmission Frequency) vary according to the smart farming system.

2.3. Designed Method

The problem in this study was tried to be solved with supervised learning techniques. Each of the data consisting of different combinations of system inputs used as features belongs to one of three different communication technology classes: WAN, LAN and LPWAN. This dataset was divided into two, one part was used for training and the other part for testing. Algorithms perform learning by seeing the class to which the data used for training belongs. Then, the algorithms are faced with new test data that is not included in the training set and are made to guess which class they belong to. The class predictions produced by the algorithms are compared with the real classes of the test data, and the performances of the algorithms are measured and compared with each other.

A MATLAB script prepared for this study was used during the multiplication and randomization of the dataset created with the scaled versions of the default values. A free ready-made program called Orange Data Mining was used during the design, testing and optimization of the created data model. This program contains ready-made machine learning algorithms. Thanks to this software, different machine learning algorithms were used on the data model. At the same time, the hyperparameters of these algorithms have been optimized.



10,000 -> 9056	11 -> 10	[0, 1]	5-fold Cross Validation

Figure 3. Data preprocessing steps

Outlier Data Removal: As a result of running the Local Outlier Factor method with 20 neighbors and Euclidean distance metric parameters; 944 data samples were found to be outliers and these data were removed from the dataset.

Feature Selection: The effect of the features determined using the Information Gain tool on the classification was observed. As seen in Table 4, the effect of the Sensor Data Transmission Frequency feature on the classification result is less than 0.01. For this reason, this feature was not used when training the machine learning algorithms.

Table 4. Information Gain Cases for Features

Feature Name	Information Gain
Data Throughput	0.178
Number of Camera Systems	0.170
Sensor Density	0.101
Land Size	0.089
Region Geography	0.086
Average Mobility of Vehicles	0.079
Reliability	0.043
Latency	0.039
Region Type	0.038
Vehicle Density	0.032
Sensor Data Transmission Frequency	0.009

Normalization: Numerical features are normalized between [0, 1].

Splitting the Dataset for Training and Testing: In the model, the 5-fold parameter was used for the layered cross-validation method, which has been accepted as successful in detecting over-learning.

The relationship between the system components and the different stages of the system for a sample application design are shown in Figure 4. Which technologies are used in which stages are also given. Different users held in the User Service object perform profile and recommendation processes through the User Interface. In case of using the recommendation system, parameters are received from the user through the User Interface. The parameters received from the user are recorded in the database at this stage. Then, the inputs received from the user are converted into inputs to be given to the machine learning model (Parameter Extraction). Through the developed Design, this data is given as input to the model and output information is received. The received model output is converted into information that users can understand through the Presentation Service and then presented to the user through the User Interface. The content of the recommendation presented to the user is recorded in the database through the User Interface.

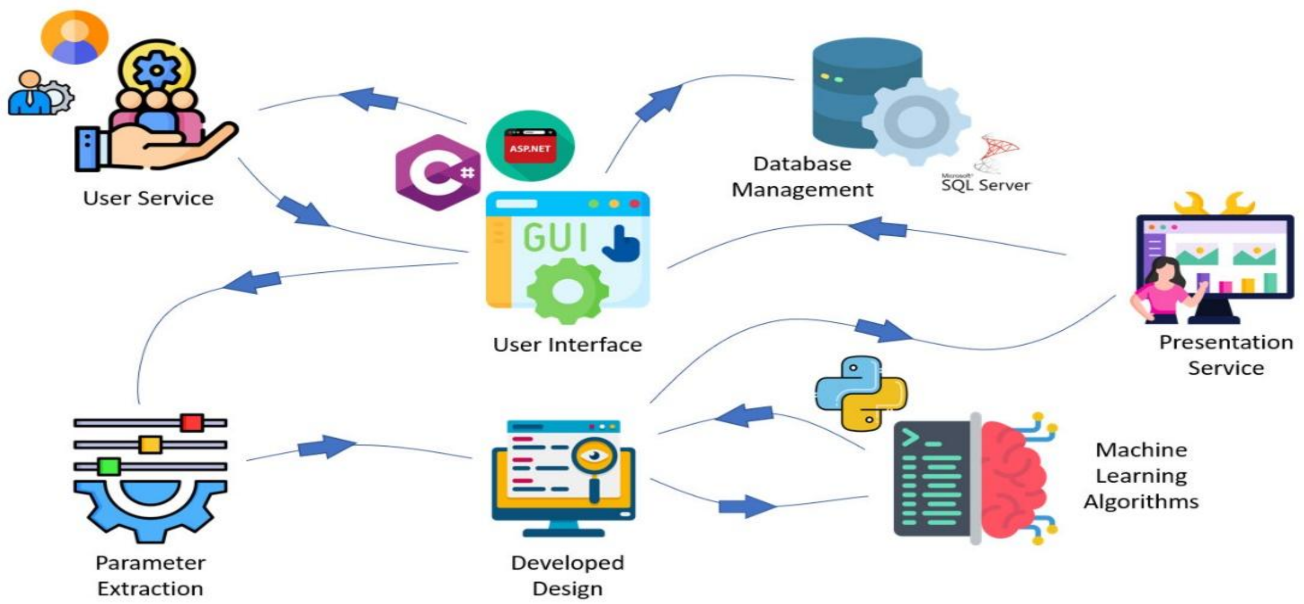


Figure 4. System architecture of the application

2.4. Case Studies

An application was developed using the system architecture given in Figure 4, and an application was created where sample case studies could be examined. In this context, different questions were directed to the user within the application, and feature values were extracted from the answers received. Three features that vary according to the terrain (Region Geography, Region Type and Land Size) were requested to be collected from the user through six different questions. These questions are listed below.

Question 1: What kind of geography is your land located in?
Plains and similar flat lands allow electromagnetic signals to travel in a direct line. Therefore, wireless signals perform relatively better in these areas. However, mountainous and wooded lands can create obstacles in the travel of electromagnetic signals, causing the signals to change direction and disperse. Therefore, signal strength may decrease or disappear completely in such rugged terrains.

Question 2: What is the relationship of your land with the water source?
Wireless communication signals can weaken when passing through high density materials such as water, which can cause some problems in wireless communication. Since the water surface is rough, the direction of the reflected signals changes and scatters at different angles. This can cause the signal strength to weaken and the signal quality to decrease. In addition, the reflected signals are also affected when moving on the surface of the water. For example, on a water surface with waves, the reflected signals can change according to the direction of the waves, which can cause the signal quality to decrease. Therefore, the signal strength may decrease or disappear completely when wireless communication devices are located near water sources.

Question 3: Is your land in a forest area?
Especially in wooded areas, signals can be blocked by trees and the signal strength can be significantly weakened. Some wireless communication technologies, especially high-frequency signals, perform better at short distances where natural obstacles are less effective.

Question 4: How many floors are the buildings around your land on average?
The height of the buildings shortens the distance that wireless signals can travel directly. Tall buildings prevent the signals from spreading directly, which can cause loss of signal strength. Tall buildings can also cause wireless signals to reflect. These reflections can cause signals to spread out at certain angles and directions, which can cause changes in signal strength.

Question 5: How sparse are the structures in your area?

The sparseness of buildings can increase the distance that wireless signals can travel directly. This means that distances between wireless communication devices can be longer, which can cause signals to weaken. In addition, in an area where buildings are sparse, wireless signals can be less reflective, which can cause signals to travel a shorter distance.

Question 6: How large is your area in acres?
A large area of land can increase the distance that wireless signals can travel. This means that distances between wireless communication devices can be longer, which can cause signals to weaken. In addition, a larger area of land can also affect the reflection of wireless signals. For example, a large plain can cause wireless signals to reflect off more surfaces. These reflections can cause signals to spread out at certain angles and directions, which can cause signals to weaken.

Each of the options created for the questions given above has been given appropriate points. A weighted score was made according to the answers to all questions and the answers were converted to system input with a value scaled between 1-10.

Parameters that change depending on smart agriculture systems (Vehicle Density, Average Mobility of Vehicles, Number of Camera Systems, Data Thorughput, Reliability, Latency, Sensor Density and Sensor Data Transmission Frequency) were accepted as a result of the analysis of predetermined weight values for these scenarios after the user was presented with different smart agriculture scenarios and made to choose from them. These scenarios are listed below.

Question 7: Which smart agriculture scenarios are intended to benefit from?
1. Harvesting and Gathering (Robot)
2. Fertilization and Irrigation (Robot)
3. Water, Mineral and Environmental Monitoring (Robot, Sensor)
4. Weed Detection (Drone + Camera)
5. Weed Control (Robot + Camera)
6. Pest Detection
7. Pest Control
8. Treatment Plans (Drone + Camera)
9. Phenotype Analysis (Drone + Robot)
10. Thinning and Pruning (Camera + Robot)
11. Sorting and Packaging (Robot + Camera)
12. Security and Lighting Applications
13. Livestock Tracking
14. Milking (Robot)
15. Sheep Herding (Drone + Camera)
16. Offspring Care

Some of these scenarios can be implemented with both sensors and camera systems, robots and drones. In order to determine the requirements of the scenarios, the user was then asked to group the scenarios he/she selected by asking three questions in this respect. It is assumed that the scenarios not marked in the questions listed below are implemented through sensors.

Question 8: In which of the scenarios you selected do you use a drone?

Question 9: In which of the scenarios you selected do you use a robot?

Question 10: In which of the scenarios you selected do you use a camera?

A sample application screen is presented in Figure 5, and the recommended result reached by the machine learning model in line with the answers to the questions is given. When this case example is examined, WAN standards are recommended according to the results given in Figure 5. It has been seen that the application working based on the developed methods can make logical decisions in different case examples. Thanks to the user-friendly questions prepared, information can be provided about what kind of wireless communication infrastructure users should prefer in the system to be preferred, without having any expertise in the wireless communication channel or wireless communication requirements. The features from the questions directed to the user are automatically extracted in the background in the application.

Terrain Conditions	Water Relationship	Forested Area	Electrical Infrastructure	Communication Infrastructure	Nearby Building Height	Land Size	Selected Scenarios	Result
High Plains and Plateaus	Near a Stream	No	No	3G	1-2	101-500	Weed Control (Robot), Water, Mineral, and Environmental Monitoring (Robot)	WAN

Figures 5. User interface for an example application

2.5. Performance Metrics

The performance tests of the developed machine learning models were performed using confusion matrix-based criteria as exemplified in Table 5. The accuracy rate was calculated using the formula (TP+TN)/(TP+TN+FP+FN). The sensitivity value was found using the formula TP/(TP+FN); and the precision value was found using the formula TP/(TP+FP). Taking both the sensitivity and precision values into account, the F1 score was calculated as (2 x Sensitivity x Precision)/(Sensitivity + Precision). While the accuracy rates show how accurately the trained model works, the F1 score value provides information about how reliable the same model is.

Table 5. Example of Confusion Matrix Used for Performance Measurement of Machine Learning Algorithms

True Positive (TP)	False Negative (FN)
False Positive (FP)	True Negative (TN)

3. Findings and Discussion

After following the preprocessing steps on the dataset created with the script prepared on the MATLAB platform, different machine learning algorithms were run with 9056 samples to determine the best hyperparameters of these algorithms. In this context, the trial and error method was followed (Yang and Shami, 2020). After determining the best hyperparameter values via the Orange Data Mining program, different machine learning algorithms were compared with each other. As a result of the experiments conducted with different parameters, the highest success values that each algorithm could achieve were collected in Table 6 for comparison. Hyperparameters for classification algorithms are also given in this table.

Table 6. Best Hyperparameters and Performance Values for Machine Learning Algorithms

Classification Algorithm	Hyperparameters	Highest Accuracy Rate	Highest F1 Score
Gradient Boosting	Method: Catboost Number of Trees: 105 Learning Rate: 0.1 Maximum Depth of Individual Tree: 4	0.842	0.841
Random Forest	Number of Trees: 25 Number of Features Considered in Each Partition: 11 Depth Limit of Individual Trees: 11	0.831	0.831
Naive Bayes	-	0.795	0.795
kNN	Metric: Manhattan Weight: Distance Number of Neighbors: 44	0.754	0.753
SVM	Cost: 1 Kernel Function: Polynomial G Value: Auto C Value: 1 D Value: 3 Numerical Tolerance: 0.001	0.741	0.740

Complexity matrices are presented separately for each classification algorithm in Figure 6. Five different machine learning algorithms were tested. Gradient boosting, random forest, naive Bayes, k-nearest neighbor value (kNN) and support vector machines (SVM) methods achieved success rates ranging from 74% to 84%. Neural networks were not included in the comparisons because they generally do not provide better results than ensemble learning algorithms on tabular datasets (Shwartz-Ziv and Armon, 2022). Among the algorithms, ensemble learning methods gave the best result, especially the highest success result of 84% was obtained with the gradient boosting classification technique. The fact that the gradient boosting method is relatively good compared to the random forest algorithm is due to the fact that the trained tree structures are corrected by looking at previous errors. The tree structures in ensemble learning methods were able to provide better results than other algorithms because they separate the data from each other by detecting the important features of the dataset. Among the classical machine learning techniques, the success results for naive Bayes, kNN and SVM remained below 80%. Classical machine learning techniques did not give good results for the tabular dataset produced.

When the complexity matrices were examined, it was seen that the most successful machine learning technique for all three classes separately was the gradient boosting algorithm. However, when the complexity matrices were examined, the random forest method, another ensemble learning algorithm, was also successful for all three classes compared to the other algorithms. It was seen that the erroneous decisions were relatively less between the WAN and LP-WAN classes, and erroneous decisions were mostly made in the separation of the LAN class. At this point, it is understood that there is a clearer separation at the feature level between the WAN and LP-WAN classes. It would not be a wrong inference that the LAN class is similar to WAN in terms of some features and LP-WAN in terms of some other features.

With the proposed system, supported by machine learning, automatic determination of which wireless communication standards will provide higher benefits by using them in a target-oriented manner is provided so that smart agricultural applications, which are becoming more and more diverse every day, can gain more functionality. The high diversity of system inputs and the consideration of important components in terms of wireless communication standards provide a significant advantage. A structure with a high level of awareness has been created due to the use of different system inputs. Another advantage is that machine learning makes decisions quickly. An application at a suitable level has also been developed for people with limited expertise.

		PREDICT		
		WAN	LAN	LP-WAN
ACTUAL	WAN	2783	144	73
	LAN	196	2587	147
	LP-WAN	111	185	2830

Gradyan Artırma

		PREDICT		
		WAN	LAN	LP-WAN
ACTUAL	WAN	2694	190	116
	LAN	235	2489	206
	LP-WAN	141	209	2786

Rastgele Orman

		PREDICT		
		WAN	LAN	LP-WAN
ACTUAL	WAN	2551	223	226
	LAN	298	2378	254
	LP-WAN	182	258	2686

Naif Bayes

		PREDICT		
		WAN	LAN	LP-WAN
ACTUAL	WAN	2423	301	276
	LAN	348	2259	323
	LP-WAN	231	311	2584

kNN

		PREDICT		
		WAN	LAN	LP-WAN
ACTUAL	WAN	2365	334	301
	LAN	374	2201	355
	LP-WAN	260	343	2523

SVM

Figure 6. Confusion matrices for machine learning algorithms

When similar studies in the literature are examined, no method that selects among different communication standards for smart agricultural applications, as in this study, has been encountered. However, the developed method will always yield better results when the performance comparison is made based on a single wireless communication standard used for smart agricultural applications. Because, the developed method aims to ensure that the most appropriate communication standard is used in that situation in a way that adapts to all conditions. Therefore, when compared to other single systems, this hybrid selection system will always be superior.

However, the operation of the algorithm that selects among different standards will bring a small additional processing load in terms of the whole system. The downside of the developed system is the need for additional data transmission and processing load. It would be realistic to talk about a potential limitation at this point. However, the developed application will not be used as a continuously operating system, but during the planning of the smart agricultural land. Therefore, the additional data transmission need and processing load brought by this system, which will be used only during planning, is not an important detail.

When the sample case study given in the previous section is examined, the questions expected to be answered from the user generally target people who do not have any knowledge about wireless communication. Thus, smart agricultural land owners and operators will be able to benefit from the developed method and application without having technical knowledge. The answers to the questions are automatically converted into attributes in the background, and the machine learning model is run in the background. The most accurate wireless communication infrastructure option is directly suggested to the user. In real-world applications, it will also be possible to suggest which smart agricultural technology products to buy in the next stage. With this feature, the developed method and application can be transformed into a complete recommendation system.

4. Conclusion

When the results of the study are examined, it is seen that the developed machine learning-based approach is promising. It has been understood that ensemble learning methods are successful in terms of tabular data sets. The original approach developed in this study can be used to decide which wireless communication standards should be used for smart agricultural systems. Both designers and users of smart agricultural systems will be able to benefit from this approach. The foundation of an important method and application for smart agricultural systems has been laid in this study.

In future studies, specific standards can be selected instead of groups of wireless communication standards. Therefore, it will be necessary to choose from more class labels. In this case, the problem may become difficult and more features may be needed. In parallel with the increase in the number of features, it may be necessary to use feature reduction techniques. Again, in subsequent studies, an application-oriented development will be made for the developed approach. At this point, various difficulties may be experienced in the process of extracting information from different systems for a real application. In addition, as another future study, smart agricultural system products can be proposed depending on the wireless communication standard.

References

Altunan, U., Sazak, H., Yazar, A. 2023. "ML-Based Service Type Priority Decision Method Using Ambient Information for 5GB", International Conference on Smart Applications, Communications and Networking (SmartNets), 1-5.

Arık, M., Korkut İ. 2022. "Irrigatıon in Agriculture and Automation Based Irrigation Systems", Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 10(2), 360-367.

Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., Aggoune, E.H.M. 2019. "Internet-of-Things (IoT)-Based Smart Agriculture: Toward Making the Fields Talk", IEEE Access, 7, 129551-129583.

Emam, K.E. 2020. "Accelerating AI with Synthetic Data", O'Reilly Media, Inc.

Eren, H. A., Adar, N., Yazar, A. 2023. "Vehicle-to-Everything Communications Standard Selection Under Different Intelligent Transportation Scenarios with Artificial Learning", Journal of Intelligent Systems: Theory and Applications, 6(1), 67-74.

Fuentealba, D., Flores, C., Soto, I., Zamorano, R., Reid, S. 2022. "Guidelines for Digital Twins in 5G Agriculture", International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), 613-618.

Hançer, A., Yazar, A. 2023a. "Waveform Decision Method with Machine Learning for 5G Uplink Communications", International Journal of Engineering Research and Development, 15(2), 820-827.

Hançer, A., Yazar, A. 2023b. "Multi-Carrier and Single-Carrier Waveform Decision Method in Non-Terrestrial Networks", Signal Processing and Communications Applications Conference, 1-4.

Hossain, M.I., Markendahl, J.I. 2021. "Comparison of LPWAN Technologies: Cost Structure and Scalability", Wireless Pers Commun, 121, 887–903.

Jape, S. D., Mungase, K. V., Thite, V. B., Jadhav, D. 2023. "A Comprehensive Analysis on 5G, IoT and Its Impact on Agriculture and Healthcare", International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 1599-1605.

Khan, M. A., Khan, A., Abuibaid, M., Huang, J. S. 2023. "Harnessing 5G Networks for Enhanced Precision Agriculture: Challenges and potential Solutions", International Conference on Smart Applications, Communications and Networking (SmartNets), 1-6.

Kihero, A.B., Tusha, A., Arslan, H. 2021. "Wireless Channel and Interference, Wireless Communication Signals: A Laboratory-based Approach", Wiley, 10, 267–324.

Liya, M.L., Arjun, D. 2020. "A Survey of LPWAN Technology in Agricultural Field", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 313-317.

Nikolenko, S.I. 2022. "Synthetic Data for Deep Learning", Springer Cham.

Ouafiq, E. M., Saadane, R., Chehri, A., Wahbi, M. 2022. "6G Enabled Smart Environments and Sustainable Cities: an Intelligent Big Data Architecture", Vehicular Technology Conference (VTC-Spring), 1-5.

Sazak, H., Yazar, A. 2023. "Ambient Aware User-Numerology Association for 5G and Beyond", Signal Processing and Communications Applications Conference (SIU), 1-4.

Shwartz-Ziv, R., Armon, A. 2022. "Tabular Data: Deep Learning is Not All You Need", Information Fusion, 81, 84–90.

Yang, L., Shami, A. 2020. "On hyperparameter optimization of machine learning algorithms: Theory and practice, Neurocomputing", 415, 295–316.

Yarkan, S., Arslan, H. 2008. "Exploiting Location Awareness toward Improved Wireless System Design in Cognitive Radio", IEEE Communications Magazine, 46(1), 128–136.

Yazar, A., Dogan-Tusha, S., Arslan, H. 2020. "6G Vision: An Ultra-Flexible Perspective", ITU Journal on Future and Evolving Technologies, 1(1), 1-20.

Yazar, A. 2021. "Requirement Analysis and Clustering Study for Possible Service Types in 6G Communications", Signal Processing and Communications Applications Conference (SIU), 1-4.

Zhang, J., Zhang, R., Yang, Q., Hu, T., Guo, K., Hong, T. 2021. "Research on Application Technology of 5G Internet of Things and Big Data in Dairy Farm", International Wireless Communications and Mobile Computing (IWCMC), 138-140.