Meta Learning-Based Dynamic Ensemble Model for Crop Selection

Bhuvaneswari Swaminathan, Saravanan Palani, and Subramaniyaswamy Vairavasundaram

School of Computing, SASTRA Deemed University, Thanjavur, India

ABSTRACT
Agricultural sector is working for optimal crop yield toward securing a sustainable food supply for the world. Fast growth in precision agriculture helps farmers to increase their yields by extending the era of machine-learning techniques. However, in organic and inorganic farming, predicting yield is an open issue that dominantly depends on the presence of soil nutrients. The lack of knowledge about the richness of land nutrients deals with the crop selection problem. Therefore, the proposed work extended the idea of a dynamic ensemble model for imbalanced multi-class nutrient data. In this work, an attempt is being made to include a novel customized voting strategy for deciding the final class output from the ensemble model. As an initial step, a well-known ranking technique, VIKOR, is applied over land nutrients to extract the most informative land samples. The rationale is to reduce the complexity of the ensemble model by determining only informative land samples for further classification. Furthermore, the meta-learning approach of dynamic ensemble selection accounts for multi-criterion-based competent classifier selection as meta-classifiers. These meta-classifiers decide on ensemble formation with the customized voting strategy to classify the right crop for the test land. To investigate nutrient richness, real-time soil and water nutrient data are collected from the soil testing laboratory, which covers different spatial data. Our experiments on six popular DES algorithms over nutrient data reveal the proposed algorithm’s outperformance in specificity, sensitivity, BCA, Multi-Area under Curve, and precision. Moreover, the lesser computational time of the proposed work indicates the model’s efficiency toward suitable crop selection.

Introduction
The rapid growth of agriculture is essential for the livelihood of human beings and the economic development of any country. However, the complex agro-ecosystem is plagued with various problems related to crop selection (Deepa and Ganesan 2019). Conventional Agriculture practices such as soil testing, fertilizer selection, and leaf disease detection can be enhanced through modern
digital technologies. As Agriculture is the backbone of the industrial sector in any country, the failure of agriculture dramatically affects the country’s GDP (Gross Domestic Product). Therefore, compulsory improvement in agricultural practices requires support to overcome many obstacles.

A vital practice called Precision Agriculture (PA) aims to develop technical awareness among farmers (Oreszczyn, Lane, and Carr 2010). Precision Farming carries out agricultural activities in a precise way that increases the crop yield without exploiting natural resources. The critical issue of any farmer is to improve their crop production, which relies on defining the suitable crop for the land. There are three significant factors that contribute to crop growth: soil, water, and season (Anilkumar et al. 2022). Soil that contains rich nutrients and excellent water retention capacity supports well during crop growth. Still, the season that is associated with climatic conditions may drastically change the expectation of crop yield at any time. Therefore, it is necessary to consider all three significant agricultural factors while providing sustainable solutions to improve crop production. Most recent researchers are also focusing on employing learning models for agricultural-related problems. Especially, the existing works that are implemented for suitable crop selection have been rely on learning models such as Johnson classifier (Deepa and Ganesan 2019), Bi-LSTM 2022 (Swaminathan et al., 2022), and advanced decision tree (Rajeswari and Suthendran 2019). However, due to the uniqueness of each land nutrient, the existing machine learning techniques lack to provide suitable crops for the land across locations. Hence, it is proven that ensemble models gained high accurate outcomes than single machine learning models (Zhu et al. 2017).

Certain emerging ensemble classifiers are Stacking (Akyol 2020), XGBoost 2021 (Gertz et al. 2020), Adaboost (Ying, Mazzuchi, and Sarkani 2020), Random Forest (RF) (Ramos et al. 2020), Gradient boosting decision tree (GBDT) (Huan et al. 2020) are static methods; in which the same ensemble classifiers are also used to label the test instances. In this case, the prediction accuracy result will not be better for each test instance. Thus, the dynamic ensemble selection strategy works based on each classifier competency (Dutta et al. 2015) to select suitable ensemble classifiers for test instances, thereby improving prediction accuracy. Figure 1 represents the two categories: Dynamic Classifier Selection (DCS) and Dynamic Ensemble Selection (DES). The DCS dynamically selects the best single classifier for each test instance (Cruz, Sabourin, and Cavalcanti 2018b; Paweland Sabourin et al. 2021), whereas DES selects a competent classifier to overcome the risk in DCS (Gao et al. 2020; García et al. 2018). Hence, using DES for the crop selection problem is a significant direction.

This work focuses on the imbalanced nature of agricultural data, which contains a majority of one regional crop for that district, and samples with other crops are significantly less in number. For this study, we have considered a north region of Thanjavur District that includes villages where paddy,
banana, and sugarcane are widely cultivated crops. Among these, most of the lands are recommended for paddy firstly; then, based on the nutrient’s richness, a few farmers will be suggested to cultivate banana and sugarcane. Due to insufficient samples for banana and sugarcane compared with paddy, any classification algorithms quickly learn the behavior of the majority class rather than the minority classes. Hence, this vital issue results in the wrong crop selection, leading to farmers’ yield loss (Deepa and Ganesan 2019). Therefore, few conventional DES techniques handle imbalanced data by including three typical stages (Pérez-Gállego et al. 2019): overproduction, where several base classifiers are taken as the pool of classifiers for training purposes; selection, where an ensemble classifier selects for each query sample and integration stage, where the output of the classifier will be combined in a way to produce single label as a final decision. Therefore, the conventional DES technique

![Figure 1. Illustration of DCS and DES techniques.](image-url)
improvement makes it suitable for handling imbalanced class problems with reduced misclassification errors.

When DES handles big data, the computational burden will be increased, which can be reduced by training only samples with high information using the ranking method. One of the popular strategies is to shrink the dataset with collective informative samples. Several multi-criteria decision-making approaches have been proposed for problems considering multiple conflicting criteria. Specifically, methods like Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), Grey Relational Analysis (GRA), Analytical Network Process (ANP), and Vlsekriterijumska Optimizacija I Kompromisno Resenje (VIKOR) are renowned Multi-Criteria Decision-Making (MCDM) techniques with the capability of choosing the best alternatives (Deepa et al. 2019) from agricultural data. Especially VIKOR is a pragmatic methodology employed for various domains such as supply chain networks, transportation cost problems, sustainable agriculture (Deepa et al. 2019), and crop selection problems (Deepa and Ganesan 2018). Therefore, the down sampling of the training dataset using the VIKOR approach before the learning model is very effective in terms of computational complexity. During the pre-processing stage of the proposed algorithm, this method attempts to solve the multi-class (Deepa et al. 2020) unbalanced issue with less complexity (Hou et al. 2020).

As the proposed methodology is applied using DES technology, the main critical stage, evaluating classifier competence for selection, should be investigated. Even though various selection criteria exist in other DES techniques, the strategy of extracting meta-features for Meta-learning is incorporated into the META-DES algorithm. It is a fact that instead of evaluating classifier competence using a single criterion, adding diversity measures among base classifiers is always beneficial for ensemble systems dynamic selection. Generally, the ensemble output defines the aggregation of each base classifier using a voting strategy. In the case of a novel proposed methodology, the dynamic voting system adopts (Zhang et al. 2019) to enhance the robustness of the classification results.

According to the initial analysis, this study attempts to develop a novel META-DES algorithm to manage imbalanced multi-label classification problems. Thus, introducing diversity measures as a new idea in META-DES and a dynamic majority voting approach improves and ensures the proposed method’s high performance. Furthermore, various experiments have been conducted using six popular DES techniques with agricultural data to prove the rightness of the proposed method. In summary, the significant contributions of this article are as follows:

1. We propose a predictive model based on a dynamic selection of ensemble classifiers for suitable crop selection that induces crop productivity.
To reduce the complexity of the ensemble model, we combine the ranking technique for extracting informative land samples from real-time nutrient data.

A novel idea of incorporating classifier diversity among a dynamic selection of ensemble classifiers ensures the formation of the most competent ensemble model.

The results of the proposed work outperform the other classification models, which could be considered a significant contribution toward the right crop selection for sustainable agriculture.

The remainder of this paper is organized as follows. Firstly, we recalled the previous related works on imbalanced classification and stages of dynamic selection in Section 2. Then, Section 3 describes the details of the novel proposed approach with equations. Next, Section 4 presents the experimental framework and results in analysis. Finally, Section 5 offers the conclusion of the paper.

**Literature Review**

In this related work section, we first elucidate the existing works on ranking strategies, then imbalanced data classification. Finally, we review the works on three stages of DES.

**Multi-Criteria Decision Making Ranking Strategies**

At present, many alternative ranking methods are commonly used, including the Technology for Order Preference by Similarity to an Ideal Solution (TOPSIS), VIKOR, Elimination and Choice Translating Reality (ELECTRE), as well as the Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), etc. (Yang et al. 2020). VIKOR is one of the widely using ranking methods in variety of applications. Among the various ranking methods, VIKOR is a much admired and well proven MCDM technique that can be applied to resolve various decision problems, including precision agriculture, business management supply chain networks, production, and design. For agriculture, the hybrid mode of VIKOR and Shannon entropy has been used to construct soft decision matrix; the authors (Deepa and Ganesan 2018), generated bijective soft rules for crop selection in agricultural lands. VIKOR has been employed in management applications, such as selecting materials (Delaram et al. 2021), identifying fiber concentrations that suit food systems (Ansarifar et al. 2019), selecting material for concrete structure repair (Kiani, Liang, and Gross 2018), choosing sustainable energy projects as part of Korea’s long-term plan (Park et al. 2021), selecting renewable energy projects (Alizadeh et al. 2020), and prioritizing land use...
conservation strategies within the reservoir basin (Lam et al. 2021). VIKOR has the following advantages over similar methods: Comparing to TOPSIS, VIKOR allows decisionmakers to decide whether a particular decision is radical or conservation based on their own needs (Lin et al. 2021). In addition, VIKOR provides higher-ranked results that are closer to the ideal solution than TOPSIS, and TOPSIS does not always provide results that are closest to the ideal solution, as the latter is not always the case (Akram, Kahraman, and Zahid 2021).

Furthermore, VIKOR takes a pragmatic approach to solving real-world issues and works well with discrete and alternative issues (Abdulkareem et al. 2020). Even though ELECTRE and VIKOR implement the same decision foundation, ELECTRE requires more computation time and has a more complex decision process (Emovon and Stephen Oghenenyenyerovwho 2020). VIKOR is a suitable choice when there are numerous alternatives and attributes to consider (Akram, Kahraman, and Zahid 2021). With PROMETHEE, if there is a significant number of criteria (more than seven), designing the problem becomes challenging, and evaluating the results is very challenging (Şahin 2021).

**Imbalanced Data Classification**

In data mining, most classification algorithms assume samples from various classes are in the same proportions. However, in reality, there are cases with one class (known as the majority class) outnumbering the remaining classes (known as the minority class). Thus, the classifier bias in favor of the majority class may lead to the misclassification of an instance, which makes the task more expensive. Such an issue is known as imbalanced classification (Wang, Minku, and Yao 2018), wherein the unbalanced number of samples belonging to inspected classes plays a vital role during classifier learning (Liu 2021). An additional problem during learning is that the number of samples from the minority class may not be enough to gain generalization, resulting in overfitting. Recent researchers have developed various approaches to tackle this difficulty that comes from the nature of data. However, many methods have been proposed, the ensemble-based classification (Kadkhodaei, Masoud Eftekhari Moghadam, and Dehghan 2020) is the center of serious research, and this is one reliable direction for data analysis (Woźniak, Graña, and Corchado 2014).

**Dynamic Ensemble Selection**

For a given unknown pattern, various classifiers behave differently based on the different classification capabilities. The method Dynamic Selection (DS) can determine a single classifier (known as Dynamic Classifier Selection, DCS) or
an ensemble of classifiers (known as Dynamic Ensemble Selection, DES) for classifying each query sample based on individual local competencies. In the DCS strategy, the local region accuracy is defined by k-nearest neighbors near the instances to be classified in the given feature space. In DES classification, the competence function \( c(\Psi, x) \) calculates the performance of classifier model \( c \) for query sample \( x \) that takes the value of interval \([0, 1]\). The ensemble \( \Psi \) represents a set of suitable classifiers chosen through combination functions such as sum, product, maximum, minimum, majority voting, fuzzy integral, and others. A schematic diagram of DCS and DES is shown in Figure 1. This DES method is extensively used in many different contexts, including imbalanced learning tasks (García et al. 2018), concept drift (Frontmatter 2004), One-versus-One (OVO) decomposition strategies (Zhang et al. 2017), and issues in classifying multi-labels (Markatopoulou, Tsoumakas, and Vlahavas 2015).

**Generation of Classifier Pool**

The goal of this stage is to build a classifier pool \( \Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_n\} \) to train the “n” number of diverse base classifiers. It is necessary to ensure the diversity among base classifiers that produce various outputs expected to improve the classification performance. This generation of diverse candidate base classifiers is a critical component in DES, and this process can be divided into two major groups (Tomasz et al. 2012). **Homogeneous ensemble**: This process involves selecting pool classifiers containing identical learning algorithms with different configurations to acquire diversity. Those identical learning algorithms can generate five main strategies, including different initializations, parameters, architectures, training datasets, and feature sets. **Heterogeneous ensemble**: In this case, the candidate classifier pool is generated using non-identical (different) based learning models, such as Neural Networks (NN), k-Nearest Neighbors (k-NN), Decision Trees (DT), and Support Vector Machines (SVM). The selection of differently structured individual classifiers generates decision boundaries well.

**Ensemble Selection**

This selection stage is the main focus of the DES technique, where the goal is to choose suitable classifiers for the query instance that need to be classified. For each query instance, it is a pivotal issue to select appropriate base classifiers upon their competence. To accomplish this goal, the ensemble selection task applied is divided into two main steps.

**Definition of the Competence Region of Query Sample**

Initially, the level of competence is estimated by a small local region in the domain feature space nearer to the query sample. So that the samples in this
region may have identical characteristics. The k-NN algorithm is typically used to determine this competence region, in which the neighbors of the query sample of \( X_j \) are selected. These neighbors/samples within the identified competence region are referred from the labeled instance known as the dynamic selection data set (DSEL). The previous DES works that are popularly available following different approaches to obtain the competence region are KNN-Equality (KNNE) (Cruz, Cavalcanti, and Ing Ren 2018), where an equal number of samples is required, KNOP, a decision-space-based method, and META-DES (Cruz et al. 2015), incorporated both feature and decision spaced approaches.

**Selection Criterion to Evaluate the Competence of the Base Classifier**

Each DES classifier possesses different criteria to assess its performance. DES-P method chooses the classifiers based on the accuracy of the competence region should be more significant than the formula (is the number of samples within the competence region). An approach KNORA-Union (KNU) chooses results from the classifier corresponding to the correct classifier samples combined by majority vote. Another method named KNOP (Cavalin, Sabourin, and Suen 2012) is the same as KNU, with the difference that KNU works based on feature space, whereas KNOP works on decision space. KNORA-Eliminate (Kwak 2008) would choose the base classifier only if all the instances in the competence region are classified correctly. Later, the META-DES methodology was proposed (Cruz et al. 2015) based on the Meta feature extraction of meta-problem with multiple criteria to measure the competence level of base classifiers. To ensure the ensemble function, the group-based criteria include interaction among candidate pool classifiers, diversity (Santos et al. 2009; Yan et al. 2021), and ambiguity (Santos, Sabourin, and Maupin 2007). More recently, the DES-MI method (García et al. 2018) has been widely used for imbalanced learning problems, where the calculation of competence is done by aggregating and weighting the local accuracy for instances in the region of competence. Accordingly, this method has good performance in classifying instances of minority classes. During selecting suitable classifiers, the foremost well-known DES techniques predominantly depend on the criteria accuracy to obtain individual classifier competence. The above-discussed DES techniques were proposed with different criteria, where the single criteria are insufficient to select the appropriate ensemble classifier subset for handling imbalanced data. The DES-MI algorithm evaluates and selects base classifiers based on the weighted accuracy rate.

It should be noted that other well-known DES algorithms employ various evaluation criteria (including individual and group-based measures). A META-
DES (Cruz, Sabourin, and Cavalcanti 2015) is a variant of DES algorithm that attempts to improve the robustness of its classification results by using multiple individual-specific criteria (such as probability, accuracy, behavior, etc.) to evaluate the base classifier’s ability to classify the test sample correctly. Furthermore, in the DES-KNN (k-nearest neighbors) (Hou et al. 2020) algorithm, group-centered criterion (referred to as diversity) is employed in conjunction with accuracy to select classifiers, enabling the ensemble of selected base classifiers to complement and work well together. To assess the competence of an ensemble system, a single criterion must be utilized (Filho et al. 2018), and diversity between the classifiers is beneficial. Thus, integrating META-DES and DES-KNN into DES-MI may improve its performance when dealing with classification problems.

**Classifier Integration Rules**

This stage aims to define integration rules for individual output from the selected ensemble classifiers. This task can be accomplished through any one of the three categories.

**Heuristic Way**

A non-trainable integration rule has been proposed to fuse the resultant class from the selected classifiers, namely, sum, product, maximum, minimum, median, decision templates, majority voting schemes, and the Dempster-Shafer approach (Kessentini, Burger, and Paquet 2015). These processes are easy to operate based on the classifier-dependent/independent nature (Zhang et al. 2019).

**Trainable Approach**

As the name serves, this approach uses the output from the selected classifier subset as input to another learning algorithm trained to obtain the best single output by combination. Several previous works proved that trainable combiners perform better using Neural Networks (NN) (Rahman and Verma 2013), multi-response linear regression, multi-layer perceptron (MLP), multinomial naive Bayes (MNB), and Mixture of Experts (ME).

**Dynamic Weighting Methods**

Depending on the competence of selected classifiers, this method assigns a higher weight to the higher competence to produce aggregated reasonable output. The recent research (Krawczyk, Woundefinedniak, and Francisco 2015) attempts to achieve the best classifier results by combining the weighting and dynamic approach for the initially selected subset of classifiers.
Proposed VIKOR-Based META-DES Algorithm Using Customized Voting Strategy (VMETA-DES-CV)

Based on the literature, it is revealed that modern researchers are more interested in solving the issues in traditional agriculture, namely, Artificial intelligence. However, most of the work in precision agriculture addresses the crop selection problem as the prominent issue, which the research has not focused on handling imbalanced agricultural datasets using efficient technologies. This article proposed the VMETA-DES-CV algorithm by combining the ranking strategy with a novel DES technique for solving crop selection problems. From a data scientist’s point of view, this crop selection issue can be viewed as a classification problem in which the data acquired from field visits are imbalanced with multi-label in nature. Therefore, the proposed solution must have high adaptability for an imbalanced dataset that guarantees improved classification results. In their earlier works, the authors proved that employing a dynamic approach toward ensemble selection of classifiers handles imbalanced classification problems well (Hou et al. 2020). Subsequently, a popular ranking approach is adopted to select the top-ranked alternatives in each crop. With these top-ranked crop factors, we can incur high classification efficiency with reduced computation time. To illustrate the process of ranking-based ensemble classification, the proposed workflow is given in Figure 2, and the proposed algorithm is provided separately for Meta-training and Classification phase (Algorithms 1 & 2).

Algorithm 1. Meta-Training phase

Input: Training samples \(train_{\lambda}\), Heterogeneous classifier pool \(\Phi = \{\varphi^1, \varphi^2, \ldots, \varphi^l\}\)

Parameters used: Nearest neighbors \(K\), Similar Output profiles \(K_o\), \(\text{att}_{i,j}\) scaling factor, threshold \(h_{\lambda}\).

Output: Meta-Classifier \(\lambda\)

1: \(\text{train}^*_{\lambda} = \emptyset\)
2: for all \(l_{\lambda}\) \(\in\) \(\text{train}_{\lambda}\) do
3: Calculate consensus measure \(H(l_{\lambda}, \Phi)\)
4: if \(H(l_{\lambda}, \Phi) < h_{\lambda}\) then
5: Find nearest neighbors \(K\) as region of competence \(\Theta_j\) for training data \(\text{train}_{\lambda}\)
6: Compute number of output profiles \(l_{\lambda}\) for training data \(\text{train}_{\lambda}\)
7: Determine the \(K_o\) similarity output profiles \(\Theta_j\) of \(l_{\lambda}\) using \(\text{train}_{\lambda}\)
8: for all \(\varphi^l\) \(\in\) \(\Phi\) do
9: \(FV_i = \text{MetaFeatureExtraction}(\Theta_j, V_j, \varphi^l, l_{\lambda})\)
10: if \(\varphi^l\) correctly classifies \(l_{\lambda}\) then
11: \(\text{att}_{i,j}\) set as 1 "\(\varphi^l\) is competent to classify \(l_{\lambda}\)"
12: else
13: \(\text{att}_{i,j}\) set as 0 "\(\varphi^l\) is competent to classify \(l_{\lambda}\)"
14: end if
15: \(\text{train}^*_{\lambda} = \text{train}^*_{\lambda} \cup FV_i\)
16: end for
17: end if
18: end for
19: Divide \(\text{train}^*_{\lambda}\) into 25:50 for validation and training, respectively.
20: Train \(\lambda\) using ensemble heterogeneous classifiers MLP, SVM, DT, and MNB.
21: return meta-classifier \(\lambda\).
Algorithm 2. Ensemble Selection using the meta-classifier \( \lambda \)

**Input:** Test dataset, Heterogeneous classifier pool \( \Phi = \{ \varphi^1, \varphi^2, \ldots, \varphi^l \} \), dynamic selection dataset \( D_{\text{sel}} \).  

**Parameters used:** Nearest neighbors \( K \), Similar Output profiles \( K_p \), \( att_{ij} \) scaling factor, threshold \( h_o \), diversity measure \( J \).  

**Output:** Class label as crop name.

1: for all \( I_{\text{test}} \) do #select competence classifiers  
2: \( \Phi = \emptyset \)  
3: Find nearest neighbors \( K \) as region of competence \( \theta_i \) for dynamic selection data.  
4: Compute number of output profiles \( I_{\text{test}} \) for testing data \( \Phi \).  
5: Find the \( K_p \) similar output profiles \( I_{\text{test}} \) of \( I_{\text{test}} \) for dynamic selection data.  
6: end for  
7: for all \( \varphi^l \Phi \) do  
8: \( FV = \text{MetaFeatureExtraction}(\Theta, I_l, \varphi^l; I_{\text{train}}) \)  
9: input \( FV \) to \( \lambda \)  
10: \( att_{ij} = \lambda(FV^l_j) \)  
11: \( \text{if } att_{ij} > 0.5 \)  
12: \( \Phi = \Phi \cup \{ \varphi^l \} \)  
13: end if  
14: end for  
15: for every \( \varphi^l \Phi \) do #diversity between \( \varphi^l \) and \( \varphi^i \)  
16: \( \text{if } i \neq k \)  
17: Compute the diversity of \( \varphi^l \): \( \text{Div}_l = \sum_j \text{Div}_l^j \)  
18: end if  
19: end for  
20: end if  
21: Rank the classifiers in \( \Phi \).  
22: Selected \( J \) diverse classifiers are added into \( \Phi_{x}^- \).  
23: else  
24: Rank the classifiers in \( \Phi \).  
25: Selected \( J \) diverse classifiers are added into \( \Phi_{x}^+ \).  
26: end if  
27: end for  
28: end for  
29: \( \Psi_{\text{META-DES-CV}} = \text{DynamicMajorityVoting}(\Phi_{x}^+) \) #Integrating output from the ensemble for each test sample  
30: return crop label \( \Psi_{\text{META-DES-CV}} \)

**Ranking Instances**

In case of the agricultural dataset contains soil and water nutrients for each land, in which the set of nutrients \( \varepsilon = \{ \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \} \) are taken as the criterion and each land \( L = \{ l_1, l_2, l_n \} \) are treated as alternatives/instances. Let \( A_{ij} (n \times m) \) be a decision matrix where the information of alternatives is expressed in terms of criteria. The implementation steps for the classical VIKOR method adopted from (Opricovic and Hshiung Tzeng 2007) are listed as follows.

Step 1: Construct decision matrix \( A_{ij} (n \times m) \) using land details, as given below

\[
A_{ij(nXm)} = A_{ij(n \times m)} = \begin{bmatrix}
  a_{00} & a_{01} & a_{02} & \cdots & a_{015} & a_{016} \\
  a_{10} & a_{11} & a_{12} & \cdots & a_{115} & a_{116} \\
  a_{20} & a_{21} & a_{22} & \cdots & a_{215} & a_{216} \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{n0} & a_{n1} & a_{n2} & \cdots & a_{n15} & a_{nm}
\end{bmatrix}
\]
Step 2: Based on the expert’s (in soil and water testing) opinion, each criterion is associated with reasonable weights ($\mu$) that denotes beneficial and non-beneficial concerning the application context.

$$\mu = \{\mu_1, \mu_2, \ldots, \mu_n\}$$

Step 3: Establish the best/positive ideal solution PIS, $PI_i^+ = (PI_1^+, PI_2^+, \ldots, PI_m^+)$ and worst/Negative Ideal Solution NIS, $NI_i^- = (NI_1^-, NI_2^-, \ldots, NI_m^-)$ values for
each soil and water nutrient with the indicator’s weight using Equations 1 and 2.

\[ PI^+ = \max\{t_{ij}\} \]  \hspace{1cm} (1)

\[ NI^- = \min\{t_{ij}\} \]  \hspace{1cm} (2)

where \( t_{ij} \) represents the beneficial and non-beneficial factors that hold original values.

Step 4: Compute the Utility measure (Group Utility-GU) and Regret measure (Individual Regret-IR) of each land using Equations 3 and 4, in which GU represents the alternative \( j \) from a positive ideal solution and IR refers to the distance of alternative \( j \) from negative ideal solution. Hence, the best alternative based on \( GU_j \) and worst alternative based on \( IR_j \).

\[ GU_j = \sum_{i=1}^{n} \varepsilon_j (PI^+ - t_{ij}) / (PI^+ - NI^-) \]  \hspace{1cm} (3)

\[ IR_j = \max_j \varepsilon_j (PI^+ - t_{ij}) / (f^*_i - NI^-) \]  \hspace{1cm} (4)

where \( \varepsilon_j \) refers to the criterion weights of alternatives \( j \) that expressed the relative significance.

Step 5: Compute the value of the VIKOR index of all alternatives \( Q_i \) is computed using Equations 5,6, and 7.

\[ G\bar{U} = \min_j GU_j, GU^- = \max_j GU_j \]  \hspace{1cm} (5)

\[ I\bar{R} = \min_j IR_j, IR^- = \max_j IR_j \]  \hspace{1cm} (6)

\[ \text{Agg } Q_i = W(GU_j - G\bar{U}) / (GU^- - G\bar{U}) + (1 - W)(IR_j - I\bar{R}) / (IR^- - I\bar{R}) \]  \hspace{1cm} (7)

\( GU^* \) and \( IR^* \) are maximum whereas \( GU^- \) and \( IR^- \) are minimum of \( GU_j \) and \( IR_j \). The coefficients \( W \) and \( 1 - W \) are denoted as the weight for group utility (\( G\bar{U} \)) and individual regret (\( I\bar{R} \)).

Step 6: Calculate the rank for each agricultural land by arranging the values of \( G\bar{U}, I\bar{R} \) and \( \text{Agg } Q_i \) in descending order.

**Novel Dynamic Ensemble Selection for Classification**

Before selecting the base classifiers for the classifier pool, the VIKOR method is incorporated to reduce the computational burden by ranking the alternatives. An approach for identifying suitable crops by training an imbalanced
multi-label dataset is the most critical problem to solve. Thus, a variant Dynamic Ensemble Selection (DES) exclusive employment helps to achieve this goal with improved accuracy. Thereupon, META-DES – a dexterousness technique is applied to select the suitable crop for the given land. The process involved in DES is explained in section 2, in which the META-DES decides the competent classifier based on extracting meta-features from classifiers. A new criterion is used to elect the most diverse classifiers as the final ensemble classifier for each test sample. Further, the ensemble output from suitable diverse meta-classifiers is determined through the dynamic majority voting for strong classification results. Thus, the proposed methodology is expressed as a novel VMETA-DES-CV approach with four stages: Pool of classifiers, Meta-Training (pseudo-code 2), Ensemble selection, and Integration.

As mentioned earlier, instances (lands) are denoted by \( L \) containing \( n \) lands that belong to paddy, banana, and sugarcane, which are treated as \( m \) classes \( (m = 3) \). For example, consider \( L_1, L_2, L_3 \ldots L_n \) is a subset of \( L \) defined as \( L = L_1 \cup L_2 \cup L_3 \cup \ldots \cup L_n \), where \( L_i \) represents the land with the same crop as class label \( w \). It is worth noting that each subset \( L_i \) will not be equal, and the dataset is imbalanced. Before starting with the novel DES classification stages, the top-ranked instances are treated as input. Then, split for training purposes, dynamic selection data set (DSEL), and testing purposes.

**Pool of Classifiers**

A set of heterogeneous base classifiers with divergent natures are chosen for pool generation. We treat Logistic Regression (LR), Stochastic Gradient Descent (SGD), K- Nearest Neighbours (KNN), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Decision Tree (DT), and Naïve Bayes (NB) as a base classifier with respective parameters for this proposed crop selection problem. At this moment, the classifier pool can be denoted as \( \Phi = \{\varphi^k\} \), where \( \varphi^k \) is assigned with the classifiers as mentioned above.

**Meta-Training Phase**

The target of this stage is to recognize the most competent classifier for identifying suitable crops. The specific process includes pick-out instances, determining meta-feature vectors, and training Meta classifiers. The process starts with the requirement of subset selection from the training dataset \( \text{train}_i \) to extract the meta-features. When land instances \( l_{x,\text{train}} \) applying to a pool of classifiers should create ambiguity (i.e., an equal chance of getting any classes) that is treated as an appropriate instance for selection. The absence of this ambiguous agreement can be calculated by the measure of consensus, \( H(l_{x,\text{train}}, \Phi) \), which is expected to be below. Let us consider the decision for \( l_{x,\text{train}} \) from \( \Phi \) be \( V(l_{x,\text{train}}) \), then the measure of consensus can be determined using Equation 8.
\[
H(l_{x,train}, \Phi) = \frac{1}{\# \text{of base classifiers}(k)} \sum_{i=1}^{k} \delta[V^i(l_{x,train}) - V(l_{x,train})]
\]  

(8)

where \(k\) denotes the number of base classifiers, \(V(l_{x,train})\) represents the output label for that instance “\(x\)” from the training dataset. Finally, a threshold \(h_c\) is maintained to select the appropriate sample when \(H(l_{x,train}, \Phi) < h_c\) this means that \(l_{x,train}\) it is ambiguous and can be taken for the meta-feature extraction process; as an initial step to meta-feature extraction, the below-mentioned metrics are to be checked for the competence region of the \(x^{th}\) sample in the training set \(l_{x,train}\).

\(\theta\): If \(\theta\) is region of feature-based competence, it is said that values of \(\theta = \{\theta_i\}\) are K-nearest neighbors of the training dataset \(l_{x,train}\) using the k-NN algorithm.

\(l_{x,train}\): The decision-based competence region wrapped as output profiles, \(\bar{l}_{x,train}\) that contains the decisions (i.e., label, \(\omega\)) yielded by each base classifier from \(\Phi\).

\(\theta\): The group of \(p\) most similar output profiles \(\theta_j = \{\bar{l}_1, \bar{l}_2, \ldots \bar{l}_p\}\) are generated from the samples belonging to the training dataset, \(train_\lambda\). In which each output profiles \(\bar{l}_j\) are linked with the class label (\(\omega^j\)) of \(l_j\).

Herewith these metrics, the meta-features (Cruz et al. 2015) can be calculated for classifiers \(\varphi^k\Phi\), in which each features associate to different criterion as defined below.

Neighbors’ hard classification \(f^{(1)}\): For each instance in \(\theta_j\), this \(f^{(1)}\) encodes the output label of \(\varphi^k\) as \(1\), if accuracy is high, otherwise set as \(0\).

Posterior probability \(f^{(2)}\): \(f^{(2)}\) generate posterior probability \(P(\omega^1|l_1)\) for all samples in \(\theta_j\) provided by \(\varphi^k\), where \(l_1\theta_j\).

Overall Local Accuracy \(f^{(3)}\): It stores the overall accuracy of the classifier \(\varphi^k\) over the competence region of \(\theta_j\) which calculates feature based.

Output Profile Classification \(f^{(4)}\): This set is based on a decision-based region that stores the decision of classifiers \(\varphi^k\). Each entry in this set represents the local accuracy for instance, in \(\theta_j\).

Classifier’s Confidence \(f^{(5)}\): This indicator calculates one value that corresponds to the distance perpendicular between the land sample and the decision region of the \(\varphi^k\)-classifier.

The above-mentioned five features solved the imbalance crop selection problem wisely. Now, obtain the meta-feature vector \(FV_{ij}\) for the training sample, \(train_\lambda\) and base classifier, \(\varphi^k\) from the below Equation 9.

\[
FV_{ij} = \left\{ f^{(1)} \cup f^{(2)} \cup f^{(3)} \cup f^{(4)} \cup f^{(5)} \right\}
\]  

(9)
Feature vector is linked to one class attribute variable \( att_{ij} \), where \( att_{ij} = \delta[V^i(l_j) - \omega^j] \). If \( att_{ij} \) Possess value 1, indicating that the base classifier \( \varphi^k \) is competent to classify the instance \( l_{x,train} \) from the training dataset. The matrix representation of \( FV_{ij} \), \( att_{ij} \) for all training land samples and base, classifier are known as meta-feature dataset \( train^i_j \). As a final step, the selected meta-classifier combination results in better performance. In our case, for given training land data, the MLP, SVM, DT, and MNB are heterogeneous base meta-classifier that outperforms well in the experiment.

**Ensemble Selection Phase**

This phase aims to select the most suitable ensemble classifiers for the given query land samples. Repeat the steps of the meta-training stage to choose the competence classifier for each query sample in \( l_{x,test} \). This phase uses a dynamic selection dataset \( dataset_{sel} \) to determine feature-based (\( \theta_j \)) and decision based (\( \theta_j \)) competence. Besides, the meta-feature vector (\( FV_{ij} \)) is extracted from the base classifiers on their performance in the competence region. Then, these meta-feature vectors are given as input to the meta-classifier. Finally, in the competence identification test, for each classifier, if \( att_{ij} > 0.5 \), it is selected for the ensemble classification (\( \Phi^j \)). Additionally, the novel idea of selecting diverse classifiers among selected ensembles (\( \Phi^j \)) is implemented based on improving classification results. The study (Filho et al. 2018) discusses the importance of DF measures for selecting diversity among the classifier. Therefore, Equation 10 explains the calculation of diversity among selected ensemble classifiers.

\[
Div_{ij} = \frac{C^{00}}{C^{11} + C^{01} + C^{10} + C^{00}}
\]

where the \( C^{ij} \) values are depicted in the contingency table (Table 1).

According to the diversity measure value the top classifiers are selected from the ensemble \( \Phi^j \) as final heterogeneous ensemble classifier \( \Phi^{j,j} \).

**Integration Phase**

The output labels from the selected ensemble competent classifiers (\( \Phi^{j,j} \)) are integrated using the enhanced voting rule for improved classification results on imbalanced crop data. The employment of majority voting is ensured by its
successful application in various ensemble selection techniques. The final prediction of the crop label for the test land sample \((test_x)\) is obtained through classifier competence-based weightage of majority voting. Here, the competence \((att_{i,j})\) calculation part for each candidate classifier is explained in the meta-training phase (Equation 11 and 12).

\[
V^{cv}(\omega_i|l_x,test) = \sum_{i=0}^{K} att_{i,j} * \Phi^{ij}(\omega_i|l_x,test) \tag{11}
\]

where \(\Phi^{ij}(\omega_i|l_x,test)\) denotes the output label for each competent classifier.

\[
\Psi_{META-DES-CV}(l_x,test) = arg \max_{i \in n} V^{cv}(\omega_i|l_x,test) \tag{12}
\]

where \(\Psi_{META-DES-CV}\) is the final output crop label for the test land samples that were taken by applying a customized voting approach \((V^{cv})\). The final output label is a recommendation for the farmer’s land.

**Experimental Results**

This section discusses the superiority of the proposed ranking-based novel DES technique using multi-AUC, \(p_{(\min)}\) and \(p_{(\avg)}\) over agricultural crop chemical properties dataset. The widely used five DES techniques, including DES-P, DES-MI, DES-KNN, KNORA-U, and META-DES, are considered for comparison with the proposed VMETA-DES-CV methodology. Additionally, to verify the effectiveness of the proposed technique on the crop selection problem, the performance metrics were compared with the other state-of-art approaches in the literature (Deepa and Ganesan 2018, 2019; Rajeswari and Suthendran 2019). Table 2 contains the python library of 6 DES algorithms for implementation.

**Experimental Protocol**

The experiments’ righteousness is assured by applying the same experimental protocol for each technology taken for comparison. The tests are carried out based on ten independent runs with average results reported to avoid the negative impact of randomness. The dataset was divided into training,
dynamic selection, and testing sets at the proportion of 2:1.5:1.5. During the meta-training process, the first 50% of the training dataset is involved in generating Meta features (\(\text{train}_A\)). Here, the crop selection issue is associated with an imbalanced multi-class problem where the different sets of classifiers are selected for the pool generation. Base classifiers with different behavior, including care selected for dynamic heterogeneous ensemble selection. The below-mentioning observations justify the heterogeneous ensemble selection of classifiers based on the previous study in literature.

- From DES-related articles (Cavalin, Sabourin, and Suen 2013) (Cruz, Cavalcanti, and Ren 2018) (Cavalin, Sabourin, and Suen 2010) (Santos and Sabourin 2011), it is demonstrated that the usage of weak classifier models for ensemble selection output better classification results, where k-NN, MLP and DT perform well.
- As mentioned by (Carlon et al. 2020) (Guerrero et al. 2012), the use of SGD and SVM handles multi-label classification for large dataset samples with better results.
- For a better understanding of independent variables that influence each class/crop, implementing the LR (Robles-Guerrero et al., 2019) approach shows accuracy in obtaining accuracy.
- NB (Cruz, Sabourin, and Cavalcanti 2015), one base classifier is incorporated to operate well on a small amount of dataset.

**Parameter setting:** With the heterogeneous classifier pool, the application of the proposed approach depends on the parameter setting such as the neighborhood size “K = 7,” representation of similar patterns as output profiles initially “kp = 1” and the ambiguity agreement called degree of a consensus set to hc = 95% together obtained the best classification results.

**Dataset Acquisition**

To achieve this work’s goal, we need a dataset with all macro and micronutrients of soil and water with respect to season. Since there is no availability of a benchmark dataset with soil nutrients and water nutrients extracted from the same land along with season information. So, we intend to collect data directly from the soil testing laboratory. Then, the soil nutrient dataset is collected from M. S. Swaminathan Research Foundation Centre, Thiruvaiyaru, Thanjavur District, Tamil Nadu. For the benefit of farmers, this research lab adapts a few villages to guide them in agriculture activities viz government schemes. Thus, to know the richness of each farmer’s land, they are recommended for soil and water test. They maintain the quantitative nutrients present in soil and water content as a dataset, which are extracted traditionally from the lands of villages. It is observed that paddy, sugarcane, and banana are
the crops widely cultivated at Thanjavur DT, Tamil Nadu, in large amounts. We considered 16 nutrients and recommended crops as input data with respect to seasons including kuruvi, samba and thalladi. The sample data in each crop is given as table in Table 3.

**Preparation of Agricultural Data for Learning Model**

According to this work, it is assumed that the proposed approach begins at the moment when one farmer queries experts regarding the selection of crops for their lands. Thus, the new test report concerning the crop selection goal removed the six columns with the farmer’s details like name, gender, phone number, land survey number, village, and district. In addition, farmer’s land details with the missing value of 60% were removed from the dataset. Secondly, we determined the highest correlated variables with each crop using soil and water nutrient data distribution. Sixteen columns/features are retained for further process based on higher distribution, represented as a graph for Soil pH value in Figures 3(a,b). Hence, it is clear that the “pH”-percentage of Hydrogen value has a greater impact on all three crops. Then, the categorical columns, namely CaCO3 from a soil test report, water type, geochemical type, CO3, SO4, and RSC from the water test report, contained values like “present” or “nil” to denote the presence of the nutrient. Those nutrients can be removed since they have not participated in crop growth. The only categorical column converted as numerical values is crop = \{paddy: 1, banana: 2, sugarcane: 3\}. Finally, the big agricultural data undergo a classical VIKOR ranking approach to shrink the dataset size. Thus, the top-ranked lands with compromise solutions are selected for the multi-label novel classification model.

**Design of VMETA-DES-CV Toward Crop Selection**

As a result of the scrutinization, the observations with 16 nutrient values with one recommended crop were conserved. Therefore, we consider 35,686 land nutrient records for 2019–2020 to show the maximum combination of nutrient proportions for experimental purposes. Among these, 20567, 8572 and 6547 records recommended paddy, sugarcane, and banana, respectively. Hence the occurrence of an imbalanced condition is confirmed. Based on the experimental setup explained in section 4, the selected samples undergo a VIKOR ranking mechanism to determine the top-ranked informative alternatives (here lands). The sub-section of section 3 describes each VIKOR step with equations, where the compromise solution is decided based on two conditions. By analyzing these results, it is easy to find the top-ranked agrilands with informative observations that support the usefulness and effectiveness of our novel DES technique.
Table 3. Sample Dataset with soil and water nutrients for three major crops. (s.C- sandy clay, c- clay).

| Soil nutrients | Water nutrients |
|----------------|-----------------|
|                | pH  | EC | OC | N  | P  | K  | texture | pH | EC | HCO$_3^-$ | Cl | Ca$^+$ | Mg$^+$ | Na$^+$ | K$^+$ | SAR | Crop   |
| pH             | 7.2 | 0.65 | 0.79 | 82 | 13.65 | 72 | s.c     | 7  | 0.7 | 4.5     | 2   | 4    | 12   | 0.81 | 0.09 | 0.47 | Paddy |
|                | 7.9 | 0.75 | 0.75 | 82 | 8.15  | 161.5 | s.c | 7  | 0.7 | 4.8     | 2   | 4.4  | 16   | 0.94 | 0.06 | 0.54 | Paddy |
|                | 7.8 | 0.55 | 0.55 | 69.2 | 8.15 | 90.5 | s.c | 7.1 | 0.67 | 4.6     | 2   | 3    | 2.8  | 0.86 | 0.04 | 0.5  | Paddy |
|                | 7.5 | 0.38 | 0.38 | 69.2 | 10.65 | 1.04 | s.c | 7.4 | 0.46 | 3.5     | 1   | 2.8  | 1.2  | 0.54 | 0.06 | 0.38 | Banana |
|                | 7.3 | 0.62 | 0.62 | 72.4 | 16.65 | 214.5 | s.c | 7.8 | 0.69 | 4.8     | 2   | 3.8  | 1.8  | 1.13 | 0.17 | 0.67 | Banana |
|                | 7.4 | 0.55 | 0.55 | 69.2 | 27.15 | 249 | s.c | 8.1 | 0.74 | 4.5     | 2   | 3.6  | 3    | 0.68 | 0.12 | 0.37 | Banana |
|                | 8   | 0.36 | 0.36 | 56.4 | 8.15  | 370 | c       | 7.1 | 3.27 | 8.7     | 22  | 8.8  | 8.8   | 14.7 | 0.38 | 4.97 | Sugarcane |
Consequently, the observed informative lands were trained using META-DES, which contains heterogeneous ensemble classifiers. As a well-known approach to dynamic ensemble selection, the imbalanced condition of a dataset is handled to predict the right crop for the test data lands. This study shows improved classification by incorporating new diversity criteria and customized ranking methods in the original META-DES technique. As crop selection problems are associated with farmers’ financial core, the outcome is assessed with many comprehensive metrics and experiments in successive sections.

**Baselines and Evaluation Metric**

**Baselines**

Stacking (Akyol 2020): The stacking approach combines the output of multiple base classifiers, which are homogeneous or heterogeneous. The superior base classifier assigns with larger weight value in the stacking ensemble for improvisation of final prediction.

XgBoost (Gertz et al. 2020): eXtreme Gradient Boosting is an improvement of base model gradient-boosted trees (Ahmed et al. 2020). It is composed of sequential series of weak learners that handle non-linear relationships.

Gradient Boosting Decision Tree (Huan et al. 2020): It is an iterative tree algorithm consisting of a series of weak models typically classification and regression trees (Ahmed et al. 2020). In each iteration, the classification and regression methods are trained from the remnant of the previous tree.

Random Forest (Ramos et al. 2020): Random forest is an ensemble of decision trees where the classification accuracy relies on each tree and is independent (Breiman 2013).

Ensemble Selection Performance (DES-P) (Hou et al. 2020): Classifiers are selected based on better performance rather than random selection. In this
method, the classifier accuracy inside the competence region is greater than the number of instances present in the competence region.

Dynamic Selection for Multiple Imbalance (DES-MI) (García et al. 2018): In DES-MI, the classifier competence is estimated by adding local accuracy of instances in the competence region and assigning minority instances with higher weights. Hence, this works well on selecting classifiers that learn well on minority instances.

Ensemble Selection of k – nearest neighbors (DES-KNN) (Kinal and Woźniak 2020): The selection of classifiers based on classifier competence and diversity brings the advantage of choosing the suitable classifiers with better classification performance.

KNORA-U (Junior et al. 2020): An ensemble of all classifiers that make at least one accurate prediction on the neighborhood side with proportional voting for weighted voting and accuracy on the neighborhood page.

KNORA-E (Muhammed Niyas and Thiyagarajan 2021): A group of classifiers that achieve perfect accuracy in the vicinity of the new example by reducing the size of the neighbors until at least one correct classifier is located.

**Evaluation Metrics**

The successful implementation of the proposed is validated by conducting different experiments. Commonly studied classification performance metrics include accuracy, F1-Score, sensitivity, specificity, and balanced accuracy classification BCA, which were recorded as an outcome from Equation (13), (14) (15) and (16) The explanation of terms used in metrics is given as follows: TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative).

Metrics formula for class’ i,’

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)
\]

\[
\text{sensitivity} = \frac{TP}{TP + FN} \quad (14)
\]

\[
\text{specificity} = \frac{TN}{TN + FP} \quad (15)
\]

\[
\text{BCA} = \text{sensitivity} + \text{specificity}/2 \quad (16)
\]

Since accuracy is often biased toward the majority class, it is not a good metric for an imbalanced dataset. Hence we avail of multi-classification analysis metrics, namely Multi-class Area Under Curve (M-AUC), Positive predictive value for minority classes or precision P_{(\text{min})}, and Average positive predictive value for all classes P_{(\text{avg})} (Li et al. 2020) as evaluation criteria to assess the performance of imbalanced crop selection.
problem. The estimation of M-AUC identifies the probability of each test point that belongs to each class. This measure, denoted by the approximately drawn member of a class, has a lower estimated class probability than the approximately drawn member. Additionally, it is worth noting that we adopt to determine the separability measure between classes. Thus, the comprehensive metric can compute using below Equation 17.

$$M - AUC = \frac{2}{cls(\text{cls} - 1)} \sum_{i<j} \hat{A}(i|j) \text{ where } \text{cls} = 3, i = 0, 1, 2. \quad (17)$$

As $P_{(\text{min})}$ and $P_{(\text{avg})}$ represent the performance of classification in a precision perspective, $P_{(\text{min})}$ expresses the positively predicted possibility for minority classes in a smaller size (Equation 18), and $P_{(\text{avg})}$ expresses the average precision for overall classes (Equation 19). The precision equation for the $i^{th}$ class is calculated below.

$$P_{(\text{min})} = \min_i \frac{tp_i}{tp_i + fp_i} \quad (18)$$

$$P_{(\text{avg})} = \frac{1}{\text{cls}} \sum_{i=0}^{2} \frac{tp_i}{tp_i + fp_i} \quad (19)$$

Thus, the obtained results from our proposed approach directly affect farmers’ yield in the way of crop cultivation based on rich nutrients. Higher metrics values express higher classification performance, in turn, gain higher crop yield.

**Experiment 1: Static Ensemble Techniques**

To prove the necessity of novel DES for suitable crop selection, the extensively used traditional static ensemble approaches, Stacking (Abdallah, Emna, and Boukadi 2022), XGBoost (Saini and Kumar Ghosh, Dhivya and Raj Vincent 2021), GBDT (Zhao et al. 2021), and RF (Dhivya and Raj Vincent, 2021), were taken for comparison. All mentioned classifier’s functions are imported from “sklearn” python. Considering this objective, we use two standard classification performance metrics, accuracy, and F1-score, to evaluate the approaches over the agricultural nutrient dataset. As mentioned in the experimental protocol, the results were taken from an average of ten runs of each algorithm given in Table 4.

We can claim that employing a traditional ensemble classifier gains lower accuracy values that denote the predicted crop choice deviates from the expert’s opinion. Therefore, to help farmers with similar expert knowledge, we moved to dynamic ensemble selection instead of static ensemble one.
**Experiment 2: Comparison with Other DES Techniques**

**Parameter Selection**

About five DES algorithms are considered for comparison, including DES-P, DES-MI, DES-KNN, KONAR-U, KNORA-E, and META-DES for parameter selection. For all DES algorithms, the parameter nearest neighbor value required for the region of competence is chosen using varying “k” values such as 1, 3, 5, and 7. The experiments were conducted with a heterogeneous pool of classifiers containing selected base classifiers (section experimental protocol). Figure 4 includes the overall BCA % for all 6 DES algorithms, in which the nearest neighbor value for the region of competence is selected based on the highest overall BCA. For example, during the cross-validation stage, if BCA is high using k = 7 for the maximum of the DES algorithm, it is also used for proposed testing-related experiments. Thus, this strategy’s parameter selection gives the algorithm a higher possibility for better classification results. This section also provides a comparison in Table 7 to justify the selection of META-DES among popular other available techniques. Obtaining the complete results using a similar heterogeneous pool of classifiers (explained in section) is the way to generate classification results that impact the META-DES algorithm selection for the proposed technique. The widely used other standard classification metrics are sensitivity, specificity, and BCA (Balanced Class Accuracy) are observed from experiments.

### Table 4. Classification performance comparison of static ensemble techniques.

| Static Ensemble Techniques | Accuracy | F1-Score |
|---------------------------|----------|----------|
| Stacking                  | 85.21    | 83.51    |
| Xgboost                   | 78.17    | 77.67    |
| GBDT                      | 71.5     | 73.37    |
| RF                        | 87.17    | 87.07    |

**Figure 4.** Comparison of six DES algorithms by varying K values using BCA value.
Table 5 values indicate that heterogeneous classifiers with META-DES achieved better results with > = 79% specificity, 80–88% sensitivity, and BCA report for each crop. As far as BCA value is concerned, the original META-DES algorithm outperformed all other DES techniques. Even though it is not possible to generalize a particular classifier to input data. Our proposed study used a collection of heterogeneous classifiers for the ensemble, which reported better performance with an enhanced META-DES algorithm for each class over the defined region of competence.

**Multi-Class Evaluation Metrics Representation**

To verify the impact of novel ideas over the original META-DES technique, the proposed performance is validated with other state-of-art DES techniques using metrics dedicated precisely to a multi-class imbalanced problem. Using a formula, the below-mentioned parameter values reveal the combined idea boosted the original META-DES approach to improved classification output (Table 6) in the context of the agricultural dataset.

**Table 5. Different classification metrics comparison using DES techniques for each crop.**

| DES Techniques | Metrics | Paddy | Banana | Sugarcane |
|----------------|---------|-------|--------|-----------|
| DES-P          | Sensitivity | 74%  | 62%  | 59%     |
|                | Specificity | 87%  | 77%  | 81%     |
|                | BCA       | 76%  | 60%  | 72%     |
| DES-MI         | Sensitivity | 76%  | 61%  | 57%     |
|                | Specificity | 81%  | 77%  | 70%     |
|                | BCA       | 78.5% | 66%  | 65%     |
| DES-KNN        | Sensitivity | 79%  | 78%  | 58%     |
|                | Specificity | 77%  | 72%  | 62%     |
|                | BCA       | 77%  | 74%  | 60%     |
| KNORA-U        | Sensitivity | 74%  | 71%  | 61%     |
|                | Specificity | 70%  | 54%  | 70%     |
|                | BCA       | 72%  | 57%  | 65%     |
| KNORA-E        | Sensitivity | 73%  | 58%  | 58%     |
|                | Specificity | 70%  | 54%  | 58%     |
|                | BCA       | 71%  | 54%  | 57%     |
| META-DES       | Sensitivity | 84%  | 88%  | 58%     |
|                | Specificity | 79%  | 84%  | 81%     |
|                | BCA       | 82%  | 84%  | 80%     |
| Proposed VMETA-DES-CV | Sensitivity | 95%  | 90%  | 89%     |
|                | Specificity | 87%  | 90%  | 80%     |
|                | BCA       | 91%  | 90%  | 80%     |

**Table 6. Different Multi-class performance metrics comparison between various DES techniques.**

| Algorithm   | M-AUC  | $p_{min}$ | $p_{avg}$ |
|-------------|--------|-----------|-----------|
| DES-P       | 72.5 ± 1.6 | 16.24 ± 0.48 | 34 ± 1.12 |
| DES-KNN     | 74 ± 1.11  | 32 ± 0.21  | 45.79 ± 0.51 |
| DES-MI      | 85.46 ± 2.04 | 17 ± 2.17  | 51 ± 0.23  |
| KNORA-U     | 78 ± 3.45  | 36 ± 0.41  | 49.2 ± 4.10 |
| KNORA-E     | 71.4 ± 0.41 | 16 ± 0.25  | 26.21 ± 3.1 |
| META-DES    | 92.97 ± 0.15 | 87.14 ± 1  | 92.25 ± 2.2 |
| VMETA-DES-CV | 97.21 ± 0.75 | 94.14 ± 2.26 | 93 ± 1.2  |
VMETA-DES-CV Vs. State-Of-Art Literature Works

As an additional investigation, the proposed test results matched the previous work on crop selection problems with different technical approaches. To prove the robustness of the proposed method, the experimental protocol was applied to all existing works for the round of ten consecutive runs. These comparison results are demonstrated in two different way for better understanding. First, Figure 5(a) with box plot considers accuracy for previous work and M-AUC value for our novel work. Each box represents the minimum and maximum value, the upper and lower quartile, and the median. It can be seen that (Deepa and Ganesan 2019) & (Deepa and Ganesan 2018) work with a similarly low value of accuracy that possesses a median of 83–89%. For the proposed illustration, the minimum M-AUC is 92%, which shows significantly better results. Second, Figure 5(b) with a bar chart demonstrates the computational time of literature work and proposed algorithm performance comparison. We can observe that the work (Rajeswari and Suthendran 2019) & (Deepa and Ganesan 2018) took a similar long execution time of a maximum of 376 s. The same experimental setup applies the novel META-DES technique without the VIKOR ranking method.

It is clear that the proposed works in less time, about 210 s, only after considering informative samples as input from the VIKOR methodology. Thus, the comparison chart reveals that the literature work failed to handle real imbalanced agricultural datasets and the enhancement of the proposed model. Therefore, we can conclude that the VMETA-DES-CV technique is a compelling idea for suitable crop selection, leading to farmers’ high profitability.

Discussion

In this work, we presented a novel VIKOR-based dynamic ensemble classifier called VMETA-DES technique for suitable crop selection. Crop management

Figure 5. (a.) & (b). Performance behavior comparison for existing approach and proposed technique.
is a crucial research area that many researchers are exploring to provide reliable solutions to real-time traditional agri-related problems. Thus, we collected a large dataset from MSSRF that contains soil nutrients concerning previously recommended crop that covers taluks in and around Thanjavur district. We are inspired about the implementation of ensemble classifier for multi-class crop classification. A new idea of incorporating diversity among classifier as a significant factor during dynamic selection ensures higher accuracy. To justify the selection of the DES model, the experiment was conducted among static ensemble techniques using performance metrics such as accuracy and F1-Score. Based on the results with lower accuracy, the work has been extended using DES techniques. Six conventional DES techniques have been considered for comparison to know the efficiency of fitting the original dataset with the model. The model that gained the highest accuracy has been chosen for implementing crop selection multi-class imbalance classification. The performance metrics like sensitivity, specificity, multi-AUC, and precision revealed the better performance of the novel proposed technique. Thus, the accurate crop selection model shows the direction to recent researchers working on developing sustainable agriculture techniques for major agricultural problems.

**Conclusion**

In this paper, the work of meta-learning-based dynamic ensemble classifier selection is employed for multi-label imbalanced classification problems to entertain crop selection issues. As the real-time soil and water sample collections create imbalance and big nutrient data, it is necessary to handle it precisely for further classification. Thus, to reduce the classification complexity, the current work utilized VIKOR, a multi-criterion ranking method, to shrink the dataset by only determining the significant informative land samples. Accordingly, toward selecting the ensemble model, the meta-features estimated the competent level of base classifiers to learn the training land samples. Next, the determination of classifier diversity among ensemble formations ensures the novelty part of the proposed work. Finally, the classification ability of the meta-learning-based dynamic ensemble model is enhanced by adopting a customized voting strategy for the final ensemble output. Experiments were conducted on imbalanced land nutrient data with three classes (crops) and compared against well-known DES techniques (technique with single selection criterion) and static conventional ensemble models. The observed experimental results have shown the superiority of the proposed VMETA-DES-CV through a multi-class determining metric, multi-AUC. The VMETA-DES-CV algorithm can justify improving classification results following five DES criteria as meta-features and a diversity-measure-based classifier selection strategy. When one criterion fails, the technique can show
improved performance with the help of other measures to perform ensemble selection. These research findings demonstrate the proposed technique’s effectiveness in helping farmers select suitable crops for their land. This work provides direction for researchers in precision agriculture to apply various DES technologies to agricultural crop-related problems. Follow-up studies can further analyze the introduction of new meta-features sets for better estimating meta-classifier competence.

Acknowledgment

The authors are grateful to the Indian Council of Social Science Research (ICSSR), New Delhi, for the financial support (No. IMPRESS/P580/278/2018-19/ICSSR) under IMPRESS Scheme. The authors express their gratitude to SASTRA Deemed University, Thanjavur, for providing the infrastructural facilities to carry out this research work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the Indian Council of Social Science Research [IMPRESS/P580/278/2018-19/ICSSR].

ORCID

Subramaniyaswamy Vairavasundaram http://orcid.org/0000-0001-5328-7672

References

Abdallah, B., R. G. Emna, and K. Boukadi. 2022. A machine learning-based approach for smart agriculture via stacking-based ensemble learning and feature selection methods. 2022 18th International Conference on Intelligent Environments, IE 2022 - Proceedings. IEEE. doi:10.1109/IE54923.2022.9826767.

Abdulkareem, K. H., A. A. Z. Nureize Arbaiy, B. B. Zaidan, O. S. Albahri, M. A. Alsalem, M. M. Salih, and M. M. Salih. 2020. A novel multi-perspective benchmarking framework for selecting image dehazing intelligent algorithms based on BWM and group VIKOR techniques. International Journal of Information Technology & Decision Making 19 (03):909–57. doi:10.1142/S0219622020500169.

Ahmed, W., Muhammad, Siddiqui, K. Muhammad, F. I. Fahad, and I. Siddiqui. 2020. Applied artificial intelligence predicting calorific value of thar lignite deposit: A comparison between back-propagation neural networks (BPNN), gradient boosting trees (GBT), and multiple linear regression (MLR). Applied Artificial Intelligence 34 (14):1124–36. doi:10.1080/08839514.2020.1824091.
Akram, M., C. Kahraman, and K. Zahid. 2021. Group decision-making based on complex spherical fuzzy VIKOR approach. Knowledge-Based Systems 216:106793. doi:10.1016/j.knosys.2021.106793.

Akyol, K. 2020. Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection. Expert Systems with Applications 148 (June):113239. doi:10.1016/J.ESWA.2020.113239. Pergamon

Alizadeh, R., L. Soltanisheh, P. D. Lund, and H. Zamanisabzi. 2020. Improving renewable energy policy planning and decision-making through a hybrid MCDM method. Energy Policy 137:111174. doi:10.1016/j.enpol.2019.111174.

Anilkumar, C., N. C. Sunitha, Harikrishna, N. B. Devate, S. Ramesh. 2022. Advances in integrated genomic selection for rapid genetic gain in crop improvement: A review. Planta 256 (5):1–20. doi:10.1007/s00425-022-03996-y. Springer Berlin Heidelberg

Ansarifar, E., F. Shahidi, M. Mohebbi, N. Ramezanian, A. Koocheki, and A. Mohamadian. 2019. Optimization of limonene microencapsulation based on native and fibril soy protein isolate by VIKOR method. LWT 115:107884. doi:10.1016/j.lwt.2019.02.071.

Breiman, L.E.O. 2013. Random forests _ enhanced reader.Pdf.

Carlon, A. G., B. Mansour Dia, L. Espath, R. Holdorf Lopez, and R. Tempone. 2020. Nesterov-aided stochastic gradient methods using laplace approximation for bayesian design optimization. Computer Methods in Applied Mechanics and Engineering 363(May):112909. Elsevier B.V. doi:10.1016/j.cma.2020.112909.

Cavalin, P. R., R. Sabourin, and C. Y. Suen. 2010. Dynamic selection of ensembles of classifiers using contextual information BT - multiple classifier systems. In International Workshop on Multiple Classifier Systems,145–54. Berlin, Heidelberg: Springer.

Cavalin, P. R., R. Sabourin, and C. Y. Suen. 2012. LoGID: An adaptive framework combining local and global incremental learning for dynamic selection of ensembles of HMMs. Pattern recognition 45: 3544–56. Pergamon. doi:10.1016/j.patcog.2012.02.034.

Cavalin, P. R., R. Sabourin, and C. Y. Suen. 2013. Dynamic selection approaches for multiple classifier systems. Neural Computing & Applications 22 (3):673–88. doi:10.1007/s00521-011-0737-9.

Cruz, R. M. O., G. D. C. Cavalcanti, and T. Ing Ren. November. 2018. A method for dynamic ensemble selection based on a filter and an adaptive distance to improve the quality of the regions of competence. doi:10.1109/ijcnn.2011.6033350.

Cruz, R. M. O., D. V. R. Oliveira, G. D. C. Cavalcanti, and R. Sabourin. September. 2018. FIRE-DES++: Enhanced online pruning of base classifiers for dynamic ensemble selection. doi:10.1016/j.patcog.2018.07.037.

Cruz, R. M. O., R. Sabourin, and G. D. C. Cavalcanti. 2015. META-DES:H: A dynamic ensemble selection technique using meta-learning and a dynamic weighting approach. International Joint Conference on Neural Networks (IJCNN) 1–8. doi:10.1109/IJCNN.2015.7280594.

Cruz, R. M. O., R. Sabourin, and G. D. C. Cavalcanti. 2018b. Dynamic classifier selection: Recent advances and perspectives. Information Fusion 41:195–216. Elsevier. doi:10.1016/j.inffus.2017.09.010.

Cruz, R. M. O., R. Sabourin, G. D. C. Cavalcanti, and T. Ing Ren. 2015. “META-DES: A dynamic ensemble selection framework using meta-learning.” Pattern Recognition 48 (5). Elsevier Ltd: 1925–35. doi:10.1016/j.patcog.2014.12.003.

Deepa, N., R. M. Durai, K. N. Senthil, K. Srinivasan, C.-Y. Chang, and A. K. Bashir. 2019. An efficient ensemble VTOPES multi-criteria decision-making model for sustainable sugarcane farms. Sustainability 11 (16):4288. doi:10.3390/su11164288.
Dee, N., and K. Ganesan. 2018. Multi-class classification using hybrid soft decision model for agriculture crop selection. Neural Computing & Applications 30 (4):1025–38. doi:10.1007/s00521-016-2749-y.

Dee, N., and K. Ganesan. 2019. Decision-making tool for crop selection for agriculture development. Neural Computing & Applications 31 (4):1215–25. doi:10.1007/s00521-017-3154-x.

Dee, N., B. P. Mohammad Zubair Khan, P. M. Durai Raj Vincent, P. Kumar Reddy Maddikunta, T. Reddy Gadekallu, and T. R. Gadekallu. 2020. Multi-class model for agriculture development using multivariate statistical method. IEEE Access 8:183749–58. doi:10.1109/ACCESS.2020.3028595.

Delaram, J., O. Fatahi Valli, M. Houshamand, and F. Ashtiani. 2021. A matching mechanism for public cloud manufacturing platforms using intuitionistic fuzzy VIKOR and deferred acceptance algorithm. International Journal of Management Science and Engineering Management 16 (2):107–22. doi:10.1080/17509653.2021.1892549. Taylor & Francis

Dhivya, E., and P. M. D. Raj Vincent. 2021. Multi-class model for agriculture development using multivariate statistical method. Journal of Ambient Intelligence and Humanized Computing 12(11 : 10009–22. Springer Berlin Heidelberg. doi:10.1007/s12652-020-02752-y.

Dutta, R., D. Smith, R. Rawnsley, G. Bishop-Hurley, J. Hills, G. Timms, and D. Henry. 2015. Dynamic cattle behavioural classification using supervised ensemble classifiers. Computers and Electronics in Agriculture 111 (February):18–28. doi:10.1016/j.compag.2014.12.002. Elsevier B.V.

Emovon, I., and O. Stephen Oghenenyerovwho. 2020. Application of MCDM method in material selection for optimal design: A review. Results in Materials 7:100115. doi:10.1016/j.rima.2020.100115.

Filho, L., S. Jose Augusto, M. P. C. Anne, and R. Hugo Nunes Santiago. 2018. Investigating the impact of selection criteria in dynamic ensemble selection methods. Expert Systems with Applications 106:141–53. Elsevier. doi:10.1016/j.eswa.2018.04.002.

Frontmatter. 2004. In combining pattern classifiers vols. i–xx. Wiley Online Books. Wiley. 10.1002/0471660264.fmatter

Gao, X., B. Ren, H. Zhang, B. Sun, J. Li, J. Xu, Y. He, and K. Li. 2020. An ensemble imbalanced classification method based on model dynamic selection driven by data partition hybrid sampling. Expert Systems with Applications 160:113660. Elsevier Ltd. doi:10.1016/j.eswa.2020.113660.

García, S., Z. Liang Zhang, A. Altalhi, S. Alshomrani, and F. Herrera. 2018. Dynamic ensemble selection for multi-class imbalanced datasets. Information Sciences 445–446(June : 22–37. Elsevier Inc. doi:10.1016/j.ins.2018.03.002.

Gertz, M., K. Große-Butenhuth, W. Junge, B. Maassen-Francke, C. Renner, H. Sparenberg, and J. Krüter. 2020. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Computers and Electronics in Agriculture 173 (June):105404. doi:10.1016/j.compag.2020.105404. Elsevier B.V.

Guerrero, J. M., G. Pajares, M. Montalvo, J. Romeo, and M. Guijarro. 2012. Support vector machines for crop/weeds identification in maize fields. Expert Systems with Applications 39 (12):11149–55. doi:10.1016/J.ESWA.2012.03.040. Pergamon

Hou, W.-H., X.-K. Wang, H.-Y. Zhang, J.-Q. Wang, and L. Li. 2020. A novel dynamic ensemble selection classifier for an imbalanced data set: An application for credit risk assessment. Knowledge-Based Systems 208:106462. doi:10.1016/j.knosys.2020.106462.

Huan, J., H. Li, M. Li, and B. Chen. 2020. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of chang
zhou fishery demonstration base, China. *Computers and Electronics in Agriculture* 175 (August):105530. doi:10.1016/j.compag.2020.105530. Elsevier B.V.

Junior, L. M., F. Maria Nardini, C. Renzo, R. Trani, and J. Antonio Macedo. 2020. A novel approach to define the local region of dynamic selection techniques in imbalanced credit scoring problems. *Expert Systems with Applications* 152:113351. Elsevier. doi:10.1016/j.eswa.2020.113351.

Kadkhodaei, H. R., A. Masoud Eftekhari Moghadam, and M. Dehghan. 2020. Hboost: A heterogeneous ensemble classifier based on the boosting method and entropy measurement. *Expert Systems with Applications* 157(November : 113482. Pergamon. doi:10.1016/J.ESWA.2020.113482.

Kessentini, Y., T. Burger, and T. Paquet. 2015. A dempster-shafer theory based combination of handwriting recognition systems with multiple rejection strategies. *Pattern recognition* 48 (2): 534–44. Elsevier Ltd. doi:10.1016/j.patcog.2014.08.010.

Kiani, B., R. Y. Liang, and J. Gross. 2018. Material selection for repair of structural concrete using VIKOR method. *Case Studies in Construction Materials* 8:489–97. Elsevier Ltd. doi:10.1016/j.cscm.2018.03.008.

Kinal, M., and M. Woźniak. 2020. Data preprocessing for Des-Knn and its application to imbalanced medical data classification. In *Asian Conference on Intelligent Information and Database Systems*, Cham, Switzerland, 589–99.

Krawczyk, B., M. Woundefindedniak, and H. Francisco. 2015. On the usefulness of one-class classifier ensembles for decomposition of multi-class problems. *Pattern Recogn* 48(12 : 3969–82. USA: Elsevier Science Inc. doi:10.1016/j.patcog.2015.06.001.

Kwak, N. 2008. Feature extraction for classification problems and its application to face recognition. *Pattern recognition* 41(5 : 1701–17. Elsevier Ltd. doi:10.1016/j.patcog.2007.10.012.

Lam, W. S., W. Hoe Lam, S. Hafizah Jaaman, and K. Fai Liew. 2021. Performance evaluation of construction companies using integrated entropy–Fuzzy VIKOR model. *Entropy* 23:3. doi:10.3390/e23030320.

Lin, M., Z. Chen, Z. Xu, X. Gou, and F. Herrera. 2021. Score function based on concentration degree for probabilistic linguistic term sets: An application to TOPSIS and VIKOR. *Information Sciences* 551:270–90. doi:10.1016/j.ins.2020.10.061.

Liu, J. 2021. A minority oversampling approach for fault detection with heterogeneous imbalanced data. *Expert Systems with Applications* 184 (December):115492. doi:10.1016/J.ESWA.2021.115492. Pergamon

Markatopoulou, F., G. Tsoumakas, and I. Vlahavas. 2015. Dynamic ensemble pruning based on multi-label classification. *Neurocomputing* 150 (PB):501–12. doi:10.1016/j.neucom.2014.07.063. Elsevier

Muhammed Niyas, K.P., and P. Thiyagarajan. 2021. "Alzheimer’s classification using dynamic ensemble of classifiers selection algorithms: A performance analysis." *Biomedical Signal Processing and Control* 68. Elsevier: 102729.

Opricovic, S., and G. Hshiung Tzeng. 2007. Extended VIKOR method in comparison with outranking methods. *European Journal of Operational Research* 178(2 : 514–29. North-Holland. doi:10.1016/j.ejor.2006.01.020.

Oreszczyn, S., A. Lane, and S. Carr. 2010. The role of networks of practice and webs of influencers on farmers’ engagement with and learning about agricultural innovations. *Journal of Rural Studies* 26(4 : 404–17. Pergamon. doi:10.1016/j.jrurstud.2010.03.003.

Park, D., H.-J. Fan, J.-J. Zhu, T. Kim, U. Myoung-Jin, S. Kim, S. Jeon, and K. Jung. 2021. Prioritization in strategic environmental assessment using fuzzy TOPSIS method with random generation for absent information in South Korea. *Sustainability* 13 (3):1458. doi:10.3390/su13031458.
Paweland Sabourin, Z., Sabourin, M. Robert, and R. Sabourin. 2021. Pre-processed dynamic classifier ensemble selection for highly imbalanced drifted data streams. *Information Fusion* 66:138–54. Elsevier. doi:10.1016/j.infus.2020.09.004.

Pérez-Gállego, P., A. Castaño, J. Ramón Quevedo, and J. J. Del Coz. 2019. Dynamic ensemble selection for quantification tasks. *Information Fusion* 45 (January):1–15. doi:10.1016/j.infus.2018.01.001. Elsevier B.V.

L., Qianmu, Y. Song, J. Zhang, and V. S. Sheng. 2020. Multiclass Imbalanced Learning with One-versus-One Decomposition and Spectral Clustering. *Expert Systems with Applications* 147. doi:10.1016/j.eswa.2019.113152.

Rahman, A., and B. Verma. 2013. Effect of ensemble classifier composition on offline cursive character recognition. *Information Processing & Management* 49 (4):852–64. doi:10.1016/j.ipm.2012.12.010. Pergamon

Rajeswari, S., and K. Suthendran. 2019. C5.0: Advanced decision tree (ADT) classification model for agricultural data analysis on cloud. *Computers and Electronics in Agriculture* 156 (June):530–39. doi:10.1016/j.compag.2018.12.013. Elsevier B.V.

Ramos, M., A. Paula, L. Prado Osco, D. Elis Garcia Furuya, W. Nunes Gonçalves, D. Cordeiro Santana, L. Pereira Ribeiro Teodoro, C. A. da Silva Junior, J. Li, F. Henrique Rojo Baio, et al. 2020. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture* 178 (November):105791. Elsevier B.V. doi:10.1016/j.compag.2020.105791.

Şahin, M. 2021. A comprehensive analysis of weighting and multicriteria methods in the context of sustainable energy. *International Journal of Environmental Science and Technology* 18 (6):1591–616. doi:10.1007/s13762-020-02922-7.

Saini, R., and S. Kumar Ghosh. 2021. Crop classification in a heterogeneous agricultural environment using ensemble classifiers and single-date sentinel-2A imagery. *Geocarto International* 36(19):2141–59. Taylor & Francis. doi:10.1080/10106049.2019.1700556.

Santos, D., M. Eulanda, R. Sabourin, and P. Maupin. 2009. Overfitting cautious selection of classifier ensembles with genetic algorithms. *Information Fusion* 10(2):150–62. Elsevier. doi:10.1016/j.infus.2008.11.003.

Santos, E. M. D., and R. Sabourin. 2011. “Classifier ensembles optimization guided by population oracle. In 2011 IEEE Congress of Evolutionary Computation (CEC), 693–98. doi:10.1109/CEC.2011.5949686.

Santos, E. M. D., R. Sabourin, and P. Maupin. 2007. Ambiguity-guided dynamic selection of ensemble of classifiers. In 2007 10th International Conference on Information Fusion, 1–8. doi:10.1109/ICIF.2007.4408123.

Swaminathan, B., S. Palani, K. Kotecha, V. Kumar, and V. S. 2022. IoT Driven Artificial Intelligence Technique for Fertilizer Recommendation Model. *IEEE Consumer Electronics Magazine* 1. doi:10.1109/MCE.2022.3151325.

Tomasz, W., M. Kurzynski, P. Podsadlo, and G. W. Stachowiak. 2012. A measure of competence based on random classification for dynamic ensemble selection. *Information Fusion* 13 (3):207–13. Elsevier. doi:10.1016/j.infus.2011.03.007.

Wang, S., L. L. Minku, and X. Yao. 2018. A systematic study of online class imbalance learning with concept drift. *IEEE Transactions on Neural Networks and Learning Systems* 29 (10):4802–21. doi:10.1109/TNNLS.2017.2771290.

Woźniak, M., M. Graña, and E. Corchado. 2014. A survey of multiple classifier systems as hybrid systems. *Information Fusion* 16 (1):3–17. doi:10.1016/j.infus.2013.04.006. Elsevier

Yang, J.-J., L. Huai-Wei, C.-S. Chao, C.-C. Shen, and C.-C. Yang. 2020. “Establishing a sustainable sports tourism evaluation framework with a hybrid multi-criteria decision-making model to explore potential sports tourism attractions in Taiwan.” *Sustainability* 12 (4). MDPI: 1673.
Yan, J., Y. Xu, Q. Cheng, S. Jiang, Q. Wang, Y. Xiao, C. Ma, J. Yan, and X. Wang. 2021. LightGBM: Accelerated genomically designed crop breeding through ensemble learning. *Genome Biology* 22 (1):1–24. doi:10.1186/s13059-021-02492-y. BioMed Central

Ying, Z., T. A. Mazzuchi, and S. Sarkani. 2020. M-AdaBoost-A based ensemble system for network intrusion detection. *Expert Systems with Applications* 162(December):113864. Pergamon. doi:10.1016/J.ESWA.2020.113864.

Zhang, Z. L., X. Gang Luo, S. Garcia, J. Fu Tang, and F. Herrera. 2017. Exploring the effectiveness of dynamic ensemble selection in the one-versus-one scheme. *Knowledge-Based Systems* 125(June):53–63. Elsevier B.V. doi:10.1016/j.knosys.2017.03.026.

Zhang, Z. L., Y. Yu Chen, J. Li, and X. Gang Luo. 2019. “A distance-based weighting framework for boosting the performance of dynamic ensemble selection.” *Information Processing and Management* 56 (4). Elsevier Ltd:1300–16. doi:10.1016/j.ipm.2019.03.009.

Zhao, L., X. Zhao, H. Zhou, X. Wang, and X. Xing. 2021. Prediction model for daily reference crop evapotranspiration based on hybrid algorithm and principal components analysis in Southwest China. *Computers and Electronics in Agriculture* 190 (August):106424. doi:10.1016/j.compag.2021.106424. Elsevier B.V.

Zhu, Y., C. Xie, G.-J. Wang, and X.-G. Yan. 2017. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China’s SME credit risk in supply Chain Finance. *Neural Computing & Applications* 28 (1):41–50. doi:10.1007/s00521-016-2304-x.