Fuzzy Hypergraph Feature Association Map for High-Dimensional Feature Selection in Agriculture and Remote Sensing

Rajashik Datta[1], Sanjan Baitalik[2], Sruti Das Choudhury[3,4*], Arup Kumar Chattopadhyay[5], Amit Kumar Das[2]

[1]Computer Science & Engineering (Artificial Intelligence), Institute of Engineering & Management, Kolkata, 700091, West Bengal, India.
[2]Computer Science & Engineering, Institute of Engineering & Management, Kolkata, 700091, West Bengal, India.
[3]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, 68588, Nebraska, United States.
[4]School of Computing, University of Nebraska-Lincoln, Lincoln, 68588, Nebraska, United States.
[5]BS Program, Indian Institute of Technology Madras, Chennai, 600036, Tamil Nadu, India.

*Corresponding author(s). E-mail(s): S.D.Choudhury@unl.edu;
Contributing authors: rajashikdatta215@gmail.com; rayan.baitalik@gmail.com; arup@study.iitm.ac.in; Amit@iem.edu.in;

Abstract

High-dimensional data in agriculture and remote sensing suffer from the curse of dimensionality, leading to poor model performance and high computational cost. Feature selection mitigates this by removing irrelevant and redundant features while preserving discriminative power. We propose FH-FAM (Fuzzy Hypergraph Feature Association Map), a novel supervised feature selection method that uses fuzzy hypergraphs to model higher-order feature interactions and uncertainty. It computes multi-way normalized mutual information for relevance, multi-way correlation for redundancy, applies sigmoidal and gamma fuzzy membership functions, constructs weighted fuzzy hypergraphs, and selects an optimized subset via maximal independent set after three-stage refinement. FH-FAM was evaluated on 15 public datasets (food/agriculture and remote sensing domains) using Random Forest classification (80:20 train-test split). Compared to FFAMFS, CFS,

DSCA, FCFB, FROT, and ABESS, FH-FAM achieved the highest mean accuracy (81.43%) and mean feature reduction (89.28%), with competitive execution time. The method excels in capturing complex multi-feature dependencies and uncertainty typical in crop classification, seed variety identification, and satellite-based land cover/environmental monitoring tasks.

Keywords: Feature selection; Fuzzy hypergraph; Multi-way mutual information; Multi-way correlation; High-dimensional datasets; Remote Sensing.

1 Introduction

High-dimensional data has become ubiquitous in domains such as bioinformatics, image processing, food and agriculture, and remote sensing, where the number of features often greatly exceeds the number of samples. This "curse of dimensionality" leads to increased computational complexity, higher risk of overfitting, and reduced model interpretability due to the presence of irrelevant, redundant, or noisy features. Feature selection is a key dimensionality reduction technique that identifies and retains the most informative subset of features while preserving the underlying data structure. Unlike feature extraction, it retains the original feature semantics, which is particularly valuable in interpretable applications such as crop variety classification, fruit quality assessment, land cover mapping, and environmental monitoring from satellite imagery.

Traditional feature selection methods, which mostly rely on pairwise relationships (e.g., mutual information or correlation), struggle to model higher-order interactions [1, 2], handle inherent uncertainty in agricultural and remote sensing data, or avoid manual subset-size tuning. To overcome these limitations, we propose FH-FAM (Fuzzy Hypergraph Feature Association Map), a supervised method that uses fuzzy hypergraphs to capture multi-way dependencies and uncertainty via multi-way normalized mutual information, multi-correlation, fuzzy memberships, and maximal independent set refinement on a weighted fuzzy hypergraph. The main contributions are:

- A novel fuzzy hypergraph framework for modeling higher-order feature interactions in high-dimensional data;
- Effective handling of uncertainty via sigmoidal and gamma fuzzy memberships applied to multi-way information-theoretic measures;
- A three-stage hypergraph refinement process that automates feature partitioning and selection without requiring a predefined subset size; and
- Superior performance on diverse agriculture (fruit/seed classification) and remote sensing (land cover, crop health) datasets, achieving the highest mean accuracy and feature reduction among compared methods.

The remainder of the paper is organized as follows: Section 2 reviews related work. Section 3 details the proposed FH-FAM method. Section 4 describes datasets, competing algorithms, and evaluation setup. Section 5 presents experimental results and analysis. Section 6 concludes with future directions.

2 Related work

Feature selection methods for high-dimensional data can be broadly categorized into graph-based, fuzzy, optimization-driven, hypergraph-based, and domain-specific approaches.

Graph-based methods model features as vertices and dependencies as edges, typically using mutual information or correlation to quantify relevance and redundancy. Early hybrid approaches combined these measures for supervised and unsupervised settings [3]. Unsupervised variants employed entropy-based graph structures suitable for unlabeled remote sensing data.

Fuzzy set theory addresses uncertainty by assigning partial membership degrees to feature relevance [4]. Fuzzy-rough sets [5] combine approximation and fuzziness to handle imprecise high-dimensional data such as spectral bands. Fuzzy enhancements to graph methods improve robustness in noisy agricultural and environmental datasets [4].

Optimization-driven techniques, including particle swarm optimization and genetic algorithms, search large combinatorial spaces. Multi-objective fuzzy swarm methods balance accuracy and subset size, while granular fuzzy neighborhood approaches integrate global and local search for large-scale agricultural data.

Hypergraphs extend graphs to model higher-order interactions among multiple features simultaneously essential for multi-spectral remote sensing bands or composite crop characteristics. Recent advances introduced fuzzy hypergraphs to represent multi-way dependencies with membership degrees. Hypergraph learning has seen rapid progress in representation, neural architectures, and applications including feature selection [6]. Recent advances extend fuzzy-rough sets with community detection on feature graphs [7] and incorporate hypergraph neural networks with sparse granules for heterogeneous data [8].

In domain-specific contexts, fuzzy-rough and ensemble methods enhance performance in bioinformatics and sentiment analysis. For agriculture and remote sensing, fuzzy and hypergraph approaches help manage biological variability, sensor noise, and multi-feature complexity in tasks such as fruit classification, crop health monitoring, and land cover analysis. While existing methods advance pairwise or optimization-based selection, they often lack integrated handling of higher-order interactions and uncertainty in a unified, automated framework. FH-FAM bridges this gap through fuzzy hypergraph construction and staged refinement.

3 Proposed method

We introduce FH-FAM (Fuzzy Hypergraph Feature Association Map), a supervised feature selection algorithm that leverages fuzzy hypergraphs to model higher-order feature interactions and uncertainty in high-dimensional data. It computes multi-way normalized mutual information for relevance, multi-way correlation for redundancy, applies fuzzy membership functions, constructs weighted fuzzy hypergraphs, and selects an optimized subset via maximal independent set after three-stage refinement. Recent fuzzy hypergraph models determine membership degrees or fuse multi-modal features [1, 2].

3.1 Algorithm steps

Consider generating all $\binom{N}{k}$ k-(feature) subsets from the original feature set $\mathcal{F}$. For each k-subset $F_i = \{f_{i1}, f_{i2}, \cdots f_{ik}\}$ from $\mathcal{F}$, where $f_{i1}, f_{i2}, \cdots f_{ik} \in \mathcal{F}$, the detailed steps are derived below.

Step 1: Computation of multi-way mutual information.

Given k random variables, $X_{i1}, X_{i2}, \cdots, X_{ik}$ associated with the k features in F_i, the joint entropy is given by Eq. 1, following Eq. A10.

$$H(X_{i1}, X_{i2}, \cdots, X_{ik}) = -\sum_{x_{i1}} \cdots \sum_{x_{ik}} p(X_{i1}, X_{i2}, \cdots, X_{ik})\ log\ p(X_{i1}, X_{i2}, \cdots, X_{ik}). \tag{1}$$

The multi-way mutual information is defined in the Eq. 2, which is derived from Eq. A13.

$$I(X_{i1}, X_{i2}, \cdots, X_{ik}) = \sum_{j=1}^{k} H_{ij} - H(X_{i1}, X_{i2}, \cdots, X_{ik}). \tag{2}$$

The normalized mutual information, $NMI(X_{i1}, X_{i2}, \cdots, X_{ik}) \in [0, 1]$ is defined in the Eq. 3 [4]. The parameter $\epsilon > 0$ will prevent division by zero. The denominator uses the minimum of the sum of individual entropies and the joint entropy to provide the tightest information-theoretic upper bound on the multi-way mutual information $I(X_1; \ldots; X_k)$, ensuring that $NMI \in [0, 1]$ in non-degenerate cases.

$$NMI(X_{i1}, X_{i2}, \cdots, X_{ik}) = \frac{I(X_{i1}, X_{i2}, \cdots, X_{ik})}{min\sum_{j=1}^{k} H(X_{ij}, H(X_{i1}, X_{i2}, \cdots, X_{ik})) + \epsilon}. \tag{3}$$

Following Eq. 3 the normalized mutual information for each subset, F_i is defined in Eq. 4.

$$MI_{NORM}(F_i) = NMI(X_{i1}, X_{i2}, \cdots, X_{ik}). \tag{4}$$

In a k-uniform hypergraph, each hyperedge is a subset of k vertices or features, F_i with its weight determined by the normalized mutual information (MI). This weight reflects the joint relevance of the selected features with respect to the target.

Step 2: Compute fuzzy relevance.

A fuzzy membership function, μ_R , referred to as fuzzy relevance (defined in Eq. 5 [4]), is derived from the Sigmoidal membership function and maps MI_{NORM} to a membership value within the interval (0,1). Here, "0" denotes irrelevance, while "1" signifies relevance, ensuring a smooth transition between these states. A higher $MI_{NORM}(F_i)$ results in $\mu_R(F_i)$ approaching "1", indicating stronger multi-way relevance.

$$\mu_R(F_i) = S(MI_{NORM}(F_i), b_1, b_2, b_3). \tag{5}$$

The Sigmoidal function in the interval $[b_1, b_3]$ is defined in [4]. The parameter $b_2 = (b_1 + b_3)/2$ is known as the crossover point, where $S(b_2; b_1, b_2, b_3) = 0.5$.

In hypergraphs, fuzzy relevance functions assign a membership value to each hyperedge, quantifying the degree of relevance between the feature group and the target class.

Step 3: Compute multi-way correlation.
The correlation coefficient, $\rho(F_i)$ quantifies the degree of overlap or redundancy in the information shared among the features in subset F_i (of size k). The multi-way correlation is computed as the average of all pairwise Pearson correlations among the features, as defined in Eq. 6 [4].

$$\rho(F_i) = \frac{1}{\binom{k}{2}} \sum_{1 \le a < b \le k} corr(X_{i_a}, Y_{i_b}). \tag{6}$$

Step 4: Compute fuzzy similarity.
The similarity between features is initially quantified using the correlation coefficient ρ, which is subsequently transformed into a fuzzy membership value μ_S through a gamma membership function. This transformation is governed by threshold parameters a_1 and a_2 [4]. Thus, the fuzzy similarity, $\mu_S(F_i)$, is defined in the Eq. 7 [4].

$$\mu_S(F_i) = \Gamma(\rho(F_i), a_1, a_2). \tag{7}$$

In the hypergraph, μ_S indicates the edge weights in terms of similarity and μ_R represents edge weights in terms of redundancy between features.
Step 5: Compute fuzzy union.
The fuzzy union μ_{union} is used to identify features that exhibit both low relevance and low redundancy. It is computed as the maximum of the fuzzy similarity μ_S and fuzzy relevance μ_R. The formal definition of μ_{union} is provided in Eq. 8 [4].

$$\mu_{union}(F_i) = max(\mu_R(F_i), \mu_S(F_i)). \tag{8}$$

In hypergraph terms, this function reflects that the hyperedge's strength is determined by the most dominant aspect - whether it is relevance or similarity.
Step 6: Compute fuzzy intersection.
The fuzzy intersection, $\mu_{intersect}$ classifies the features that are high both on relevance and redundancy, by taking the minimum of μ_S and μ_R. $\mu_{intersect}$ is defined in Eq. 9 [4].

$$\mu_{intersect}(F_i) = min(\mu_R(F_i), \mu_S(F_i)). \tag{9}$$

Step 7: Compute fuzzy hyper feature association.
The fuzzy hyper feature association is calculated by taking the Hadamard product ($\circ$), which is then derived by the element-wise product of μ_{union} and $\mu_{intersect}$, defined in Eq. 10 [4].

$$\mu_{FH-FAM}(F_i) = \mu_{union}(F_i) \circ \mu_{intersect}(F_i). \tag{10}$$

μ_{FH-FAM} analyses how strongly the relevance and the similarity contribute to the weight of hyperedges. It captures both multi-way relevance and correlation-based redundancy.
Step 8: Aggregate per-feature fuzzy scores.
For each feature f_i, two aggregate values are calculated respectively - $sum_{union}(f_i)$ and $sum_{intersect}(f_i)$, defined in Eq. 11 and Eq. 12 respectively [4]. The overall

relevance in union and intersection of the features is represented through these aggregated values, across all subsets F_i that contain f_i.

$$sum_{union}(f_i) = \frac{1}{|\{F_i : f_i \in F_i\}|} \sum_{F_i \ni F_i} \mu_{union}(F_i). \tag{11}$$

$$sum_{intersect}(f_i) = \frac{1}{|\{F_i : f_i \in F_i\}|} \sum_{F_i \ni F_i} \mu_{intersect}(F_i). \tag{12}$$

where the condition $F_i \in f_i$ denotes that the summation is taken over all k-subsets F_i that contain the feature f_i. The global mean of the fuzzy union across all features is calculated in Eq. 13 [4] to establish a threshold, which will be used for distinguishing features of relevance.

$$mean_{union} = \frac{1}{N} \sum_{i=1}^{N} sum_{union}(f_i). \tag{13}$$

Step 9: Partitioning the features into color groups.
Based on the computed fuzzy scores, the feature set $\mathcal{F}$ is partitioned into three disjoint groups.

F_r is the red-colored group representing features with fuzzy union value less than the global mean, indicating low relevance, as defined in Eq. 14 [4].

$$F_r = \{f_i \in F : sum_{union}(f_i) < mean_{union}\}. \tag{14}$$

F_y is the yellow-colored group representing features with fuzzy union value greater than the global mean, defined in Eq. 15 [4].

$$F_y = \{f_i \in F : sum_{union}(f_i) \geq mean_{union}\}. \tag{15}$$

F_b is the blue-colored group representing features with high redundancy, defined in Eq. 16 [4].

$$F_b = \{f_i \in F : sum_{intersect}(f_i) > mean_{union}\}. \tag{16}$$

Step 10: Hypergraph construction.
A hypergraph is constructed where the features are represented by the nodes and the higher order relationships among features are represented by hyperedges. Each hyperedge is weighted by the fuzzy parameters computed in the Steps 1 - 6. The hypergraph is able to capture the multi-feature interaction which enables the modelling of complex relevance and redundancy patterns in the data.

Let X be the set of accepted features after filtering. Thus, for all k-subsets $\mathcal{F} \subseteq \zeta$, hypergraph is defined in the Eq. 17. We denote by $\zeta_k(f_j)$ the aggregated fuzzy association score for feature f_j with respect to k-subsets, obtained by averaging (or summing) $\mu_{FH-FAM}(F_i)$ over all k-subsets F_i that contain f_j. In Fig. 1 (Left) the hypergraph construction is shown before coloring.

$$H = (X, \zeta_k) \text{ where } \zeta_k = \mu_{FH-FAM}(\mathcal{F}). \tag{17}$$

As part of this construction, the labels of r (red), y (yellow) and b (blue) are assigned based on the group they belong to.

Step 11: Three-stage hypergraph refinement and final feature selection. The threshold τ, established in Eq. 18 [4], is computed as the average fuzzy relevance score over all $N-1$ other features (excluding the feature itself in a leave-one-out sense), providing a data-adaptive, feature-specific reference for distinguishing weakly relevant features.

$$\tau = \frac{1}{N-1} \sum_{f_i \in F} \mu_R(F_i). \tag{18}$$

The complete procedure is detailed in Algorithm 3.1.

3.2 Illustration of stages

The stage-wise feature selection using a sample hypergraph is illustrated below:

Stage 1: Features with relevance below the threshold τ are moved to F_k. Fig. 1 (Right) shows Stage 1 of FH-FAM, where 4 features are already marked "red" (F_r) for direct rejection, 2 features are "yellow" (F_y) (to be reassigned later as "red" or "green"), and "blue" features (F_b) exhibit both high relevance and high redundancy.

Stage 2: In Fig. 1 (Left), "yellow" vertices are reassigned as "red" or "green" per the partitioning rules in Step 9. "Red" features have low relevance and high similarity, while "green" features have high relevance and low similarity. "Green" vertices are retained as final selected features; "red" vertices are rejected. The "green" set is defined as

$$F_g = F_y \bigcap F_k. \tag{19}$$

Stage 3: Fig. 2 (Right) shows the final FH-FAM configuration. A subset of the originally "blue" vertices (F_b, per Eq. 16) is selected via fuzzy maximal independent set. Of the three "blue" vertices in Stage 1, one is reassigned to "green" and the rest to "red". Ultimately, "green" vertices represent the selected features, while "red" vertices are rejected.

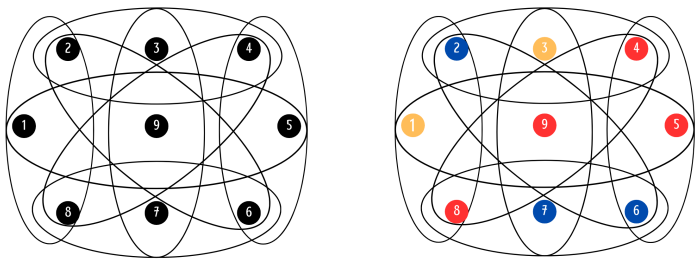

Fig. 1 Stages 1 (left) and 2 (right) of FH-FAM.

Algorithm 1 Fuzzy Hypergraph Feature Association Map (FH-FAM)

Input: Dataset D_N with N-dimensions i.e. having feature set $\mathcal{F}$ (where $\mathcal{F} = \{f_1, f_2, f_3, \cdots, f_N\}$)
Output: Optimal feature subset, $\mathcal{F}_{opt}$
Notations used: For every k-subset F_i from $\mathcal{F}$, $F_i = \{f_{i1}, f_{i2}, \cdots, f_{i_k}\}$, where $f_{i1}, f_{i2}, \cdots, f_{i_k} \in \mathcal{F}$,
b_1, b_2 and b_3 are thresholds of Sigmoidal membership function,
a_1, and a_2 are thresholds of Gamma membership function,
F_r, F_b, F_y, F_g - features marked "red", "blue", "yellow" and "green" respectively
$G_{FH-FAM1}$ and $G_{FH-FAM2}$ are the k-uniform fuzzy hypergraphs.

Begin
for each $F_i \in \mathcal{F}$, **do**
  $MI_{NORM}(F_i) = MNI(f_{i1}, f_{i2}, \cdots, f_{ik})$
  $\mu_R(F_i) = S(MI_{NORM}(F_i), b_1, b_2, b_3)$
end for
for each $F_i \in \mathcal{F}$, **do**
  $\rho(F_i) = MultiCorrelation(f_{i1}, f_{i2}, \cdots, f_{ik})$
  $\mu_S(F_i) = \gamma(\rho(F_i), a_1, a_2)$
end for
for each $F_i \in \mathcal{F}$, **do**
  $\mu_{union}(F_i) = max(\mu_R(F_i), \mu_S(F_i))$
  $\mu_{intersect}(F_i) = min(\mu_R(F_i), \mu_S(F_i))$
  $\mu_{FH-FAM}(F_i) = \mu_{union}(F_i) \circ \mu_{intersect}(F_i)$
end for
for each $F_i \in \mathcal{F}$, **do**
  $sum_{union}(f_i) = mean(\mu_{union}(F_i))$ where $\{F_i | f_i \in F_i\}$
  $sum_{intersect}(f_i) = mean(\mu_{intersect}(F_i))$ where $\{F_i | f_i \in F_i\}$
end for
$mean_{union} = mean(sum_{union}(\{f_i\}_{i=1}^{N}))$
for each $f_i \in \mathcal{F}$, **do**
$F_r \leftarrow \{f_i \in \mathcal{F} | sum_{union} < mean_{union}\}$
$F_y \leftarrow \{f_i \in \mathcal{F} | sum_{union} \geq mean_{union}\}$
$F_b \leftarrow \{f_i \in \mathcal{F} | sum_{intersect} > mean_{union}\}$
end for
$G_{FH-FAM1} \leftarrow GenerateHypergraph(\mu_{FH-FAM}, F_r, F_b, F_y)$
$\tau \leftarrow \frac{1}{N-1} \sum_{f_i \in \mathcal{F}} \mu_R(F_i)$
$F_k \leftarrow \{f_i \in \mathcal{F} \mid \mu_R(F_i) < \tau\}$
$F_g \leftarrow F_y \cap F_k$
$F_r \leftarrow F_r \cup (F_y \setminus F_g)$
$F_y \leftarrow F_y \setminus F_g$
$G_{FH-FAM2} \leftarrow UpdateHypergraph(G_{FH-FAM1}, F_r, F_b, F_g)$
$F_b' \leftarrow findMIS(G_{FH-FAM2}, F_b)$
$F_g \leftarrow F_g \cup F_b'$
$F_r \leftarrow F_r \cup (F_b \setminus F_b')$
$F_b \leftarrow F_b'$
$G_{FH-FAM3} \leftarrow UpdateHypergraph(G_{FH-FAM2}, F_r, F_b, F_g)$
Return $\mathcal{F}_{opt} \leftarrow F_g$
End

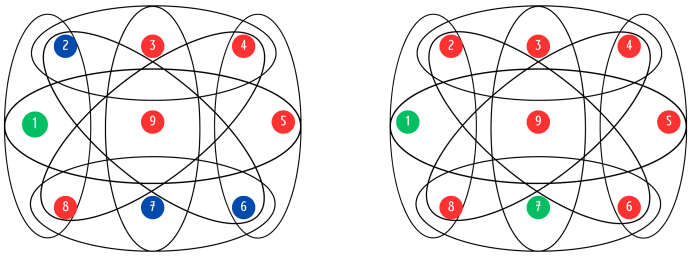

Fig. 2 Stages 3 (left) and 4 (right) of FH-FAM.

4 Methods and materials

4.1 Datasets

FH-FAM was evaluated on 15 publicly available datasets spanning food/agriculture and remote sensing domains. These datasets vary in dimensionality (7-236 features), sample size (523-1,048,575 records), and classification complexity (binary to 7 classes). Key characteristics are summarized in Table 1. All datasets were preprocessed with an 80:20 train-test split (stratified where applicable). The Random Forest classifier served as the downstream evaluation model for consistency across experiments.

Datasets	Number of Features	Number of Records	Number of Classes
Date Fruit [9]	34	898	7
Dry Bean [10]	16	13611	7
Durum Wheat [11]	236	9000	3
Pistachio 16 features [12]	16	2148	2
Pistachio 28 features [12]	28	2148	2
Pumpkin Seeds [13]	12	2500	2
Raisin [14]	7	900	2
Rice MSC [15]	106	75000	5
Coverttype [16]	54	581012	7
Winnipeg Dataset [17]	174	325834	7
crop health [18]	31	212019	2
forest type mapping [19]	27	523	4
landsat data [20]	36	6435	7
time series [21]	5	1048575	7
weather classification data [22]	10	13200	4

Table 1 Datasets used for experimentation.

4.2 Competing algorithms

Performance was compared against the following supervised feature selection methods. All methods were executed under identical hardware/software conditions and hyperparameters (default where not specified).

1. FFAMFS [4]: Fuzzy graph-based association mapping (baseline for fuzzy approaches).
2. CFS [23]: Correlation-based subset evaluation with best-first search.
3. DSCA [24]: Dominant set clustering on feature similarity graphs.
4. FCFB [25]: Fast filter using symmetrical uncertainty and approximate Markov blankets.
5. FROT [26]: Forward rotational greedy selection based on model performance.
6. ABESS [27]: Adaptive best subset selection via polynomial-time sequencing.

4.3 Evaluation criteria

Three metrics were used:

- Classification accuracy: $\frac{TP+TN}{TP+TN+FP+FN}$ on the test set.
- Feature reduction (%): $\left(1 - \frac{|\mathcal{F}_{opt}|}{N}\right) \times 100$.
- Execution time: Wall-clock time (seconds) for feature selection.

Mean values across datasets and Win/Draw/Loss counts (relative to FH-FAM) are reported for comparison.

5 Results and analysis

FH-FAM was evaluated on 15 public datasets from food/agriculture and remote sensing domains using Random Forest classification (80:20 stratified train-test split). Random Forest was selected as the downstream classifier because of its well-established robustness to high-dimensional data, inherent ability to handle non-linear feature interactions and mixed data types, resistance to overfitting through ensemble averaging and feature subsampling, and minimal sensitivity to hyperparameter tuning—properties that make it particularly suitable as a stable and reproducible evaluation model for comparing feature selection methods across diverse datasets.

5.1 Accuracy and feature selection

The comparative accuracy results of the proposed FH-FAM algorithm against competing methods (FCFB, CFS, DSCA, FFAMFS, FROT, ABESS) are presented in Table 2 and visualized in Fig. 3. Table 2 shows that FH-FAM ranks second-best overall and achieves the highest mean accuracy (81.43%) across the 15 datasets. A "Win" indicates FH-FAM attained the strictly highest accuracy on a dataset; a "Loss" means at least one competitor outperformed it; no exact ties occurred.

From Table 2 and Fig. 3, FH-FAM records the highest accuracy on 5 of 15 datasets and the best mean accuracy of **81.43%**, surpassing all competitors on average. It achieves the second-highest number of wins (5/0/10 Win/Draw/Loss), trailing only ABESS (6/0/9), while showing a strong balance of wins and fewer outright losses. These results demonstrate FH-FAM's consistent robustness in capturing complex feature interactions in high-dimensional agricultural and remote sensing classification tasks.

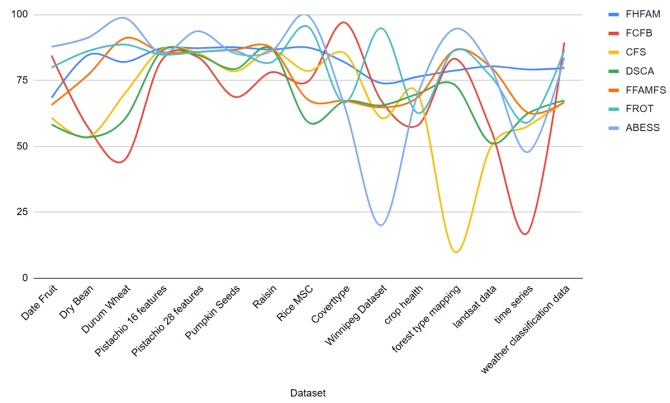

Fig. 3 Comparison of classification accuracy (%) and competing feature selection methods.

Dataset	FH-FAM	FCFB	CFS	DSCA	FFAMFS	FROT	ABESS
Date Fruit	68.37	84.41	60.9	58.27	65.7	79.85	**87.78**
Dry Bean	84.62	57.11	53.5	53.4	77.15	86.13	**91.25**
Durum Wheat	82	44.86	70	60.33	91	88.6	**98.69**
Pistachio 16 features	**87.22**	82.49	86.7	86	85.84	84.8	85.34
Pistachio 28 features	87.22	83.8	85.38	84.6	85.84	85.75	**93.7**
Pumpkin Seeds	**87.6**	68.72	85.45	79.3	86.66	86.26	85.3
Raisin	86.67	78.11	86.29	87.03	**87.4**	81.85	86.11
Rice MSC	87.5	74.56	85.34	59.21	67.7	95.4	**99.7**
Covertype	81.89	**96.98**	85.34	67.96	64.78	66.57	65.3
Winnipeg Dataset	74	67.39	60.79	65.49	64.78	**94.8**	20
Crop Health	**76.4**	57.93	69.99	69.99	68.6	62.6	69.85
Forest Type Mapping	78.7	83.94	10	73.37	86.36	86.53	**94.5**
Landsat Data	**80.37**	56.22	49.94	51.2	80	76.53	79.67
Time Series	**79.15**	17.2	57.6	62.43	62.8	59	47.75
Weather Classification Data	79.69	**89.38**	56.7	67.42	66.74	86.46	83.65
Mean	**81.43**	69.49	66.69	68.33	76.26	81.41	79.24
Win/Draw/Loss	5/0/10	2/0/13	0/0/15	0/0/15	1/0/14	1/0/14	6/0/9

Table 2 Accuracy performance for FH-FAM algorithm.

In the *Food and Agriculture* domain, involving morphological and textural features for fruit, seed, and crop classification, FH-FAM delivers superior accuracy on datasets such as Pistachio and Pumpkin Seeds. This stems from its fuzzy hypergraph structure, which effectively models uncertain and higher-order feature relationships—common in agricultural data due to environmental variability and sample quality differences—outperforming FFAMFS (fuzzy graphs lacking hypergraph multi-feature support) and ABESS (sparse regression), thus yielding more robust feature subsets for crop quality assessment. In the *Remote Sensing* domain, with high-dimensional spectral and temporal data for land cover and environmental classification, FH-FAM excels on Crop Health, Landsat Data, and Time Series datasets. Its hypergraph representation captures complex multi-band correlations and preserves higher-order relations, surpassing graph-based methods like DSCA

and FFAMFS, while the fuzzy component handles noise and uncertainty in satellite imagery—delivering higher accuracy than ABESS on large-scale data and suiting applications such as forest mapping and weather monitoring.

5.2 Percentage of feature reduction

Table 3 shows a comparison of the feature reduction of FH-FAM with the competing algorithms. Fig. 4 demonstrates the graphical comparison of the percentage of feature reduction. From the results in Table 3 and Figure 4, FH-FAM demonstrates strong feature reduction performance compared to the competing methods (FCFB, CFS, DSCA, FFAMFS, FROT, ABESS). It achieves the highest reduction on 2 of the 15 datasets, ties for the highest on 5 more, and attains the best overall mean feature reduction of **89.28%**, clearly outperforming all baselines in average capability. FH-FAM records a Win/Draw/Loss count of 2/5/8, ranking second in strict wins while showing one of the lowest numbers of outright losses. These trends highlight FH-FAM's superior ability to eliminate redundant and irrelevant features while preserving discriminative power—especially valuable for high-dimensional agriculture and remote sensing datasets where computational efficiency is critical.

Although high reduction rates (e.g., 98.73% on Durum Wheat) greatly improve efficiency, they carry a theoretical risk of information loss in tasks sensitive to subtle discriminative features; our results indicate this risk is low, as accuracies remain competitive or superior, though task-specific sensitivity analysis is advised for critical applications. In *Remote Sensing*, FH-FAM's superior mean reduction arises from its hypergraph-based modeling of complex spectral relationships, outperforming correlation-based CFS and graph-based DSCA. Compared to recent evolutionary or wrapper methods in remote sensing, FH-FAM delivers stable, effective dimensionality reduction, enhancing scalability for large-scale environmental monitoring.

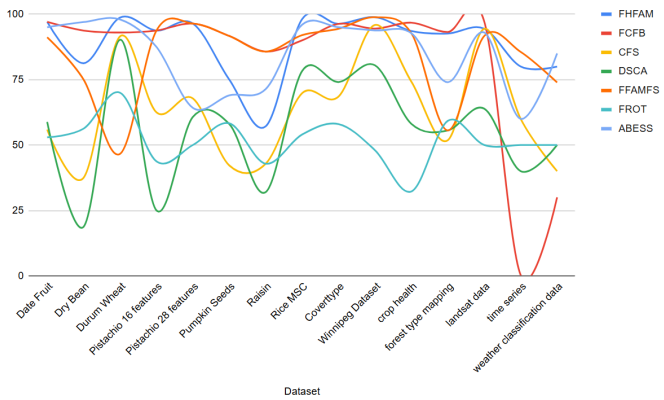

Fig. 4 Comparison of feature reduction (%) and competing feature selection methods.

Dataset	FH-FAM	FCFB	CFS	DSCA	FFAMFS	FROT	ABESS
Date Fruit	**97.06**	**97.06**	55.88	58.82	91.18	52.94	95
Dry Bean	81.25	93.75	37.5	18.75	75	56.25	**97**
Durum Wheat	**98.73**	93	91.52	90.25	46.61	70	98
Pistachio 16 features	**93.75**	**93.75**	62.5	25	**93.75**	43.75	87.5
Pistachio 28 features	**96.43**	**96.43**	67.8	60.71	**96.43**	50	64.2
Pumpkin Seeds	75	**91.67**	42.18	58	**91.67**	58.3	69
Raisin	57.14	**85.71**	42.85	32	**85.71**	42.85	71.4
Rice MSC	**98.11**	90	69.81	78.3	92	54	96
Covertype	**96.3**	**96.3**	68.5	74.07	94.44	58	95
Winnipeg Dataset	**98.85**	94.5	95.9	80.46	**98.85**	48	93.8
crop health	93.55	**96.77**	74.19	58.06	92.7	32.25	92.7
forest type mapping	92.59	**93.2**	51.85	55.56	55.56	59.25	74
landsat data	94.44	**97.22**	94.44	63.89	91.67	50	60
time series	80	0.2535	60	40	**85.6**	50	60
weather classification data	80	30	40	51.6	73.9	50	**85**
Mean	**88.88**	83.31	63.66	56.26	84.34	51.71	84.77
Win/Draw/Loss	2/5/8	3/6/6	0/0/15	0/0/15	1/5/9	0/0/15	2/0/13

Table 3 Feature reduction for proposed FH-FAM algorithm.

5.3 Execution time

From Table 4, and its graphical representation in Fig. 5, it is inferred that the proposed FH-FAM algorithm performs at Rank 5 in case of model execution time, as compared to the 7 competing algorithms. The mean execution time of FH-FAM is 11.08, representing average performance in this criteria. While FH-FAM's execution

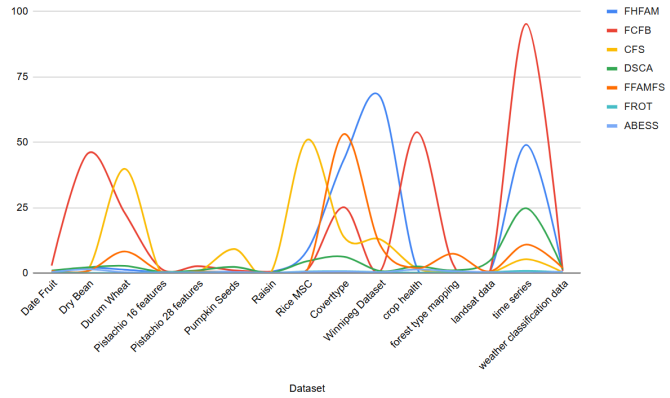

Fig. 5 Comparison of execution time (seconds) and competing feature selection methods.

time is average overall, it scales reasonably in both domains. In *Food and Agriculture*, the fuzzy handling of uncertainty adds computational overhead but is justified by accuracy gains. In *Remote Sensing*, for large datasets like Covertype and Time Series, FH-FAM is slower than ABESS but offers better trade-offs in accuracy and reduction, aligning with needs for efficient processing in real-time monitoring applications.

Dataset	FH-FAM	FCFB	CFS	DSCA	FFAMFS	FROT	ABESS
Date Fruit	0.25	2.8	0.32	0.923	0.771	0.303	0.323
Dry Bean	2.02	45.79	1.355	2.15	0.787	0.026	1.393
Durum Wheat	1.23	23	39.83	2.722	8.23	0.052	0.246
Pistachio 16 features	0.35	1.66	0.128	0.44	0.564	0.305	0.117
Pistachio 28 features	0.37	2.58	0.35	1.004	0.523	0.29	0.118
Pumpkin Seeds	0.48	0.9686	9.15	2.305	0.242	0.26	0.349
Raisin	0.31	0.47	0.032	0.111	0.169	0.033	0.11
Rice MSC	8.59	1.23	50.95	4.48	1.088	0.58	0.417
Covertype	43.23	25.14	13.87	6.233	53.1	0.687	0.56
Winnipeg Dataset	67.38	0.6752	12.85	0.712	10.64	0.33	0.322
crop health	2.18	53.84	1.82	2.49	1.85	0.14	1.3
forest type mapping	0.34	3.26	0.297	1.008	7.34	0.298	0.689
landsat data	0.49	0.34	0.04	4.63	0.589	0.341	0.279
time series	48.99	95.24	5.26	24.76	10.85	0.79	0.06
weather classification data	0.83	0.8439	0.22	1.39	1.89	0.269	0.496
Mean	11.80	17.19	9.1	3.69	6.58	**0.31**	0.45
Win/Draw/Loss	1/0/14	0/0/15	4/0/11	0/0/15	1/0/14	3/0/12	6/0/9

Table 4 Execution time for proposed FH-FAM algorithm.

5.4 Statistical significance analysis

To determine whether the observed performance differences between FH-FAM and the competing methods are statistically significant, we conducted pairwise Wilcoxon signed-rank tests on the classification accuracy, feature reduction percentage, and execution time metrics across the 15 datasets. The Wilcoxon signed-rank test is a non-parametric statistical method suitable for paired comparisons, testing the null hypothesis that the median difference between FH-FAM and each competitor is zero. We used a significance level of $\alpha = 0.05$ and considered the direction of improvement (higher accuracy and feature reduction; lower execution time). The p-values for each comparison are summarized in Table 5.

Metric	vs. FCFB	vs. CFS	vs. DSCA	vs. FFAMFS	vs. FROT	vs. ABESS
Accuracy	0.048*	0.0004*	0.0001*	0.022*	0.890	0.890
Feature Reduction (%)	0.790	0.001*	<0.001*	0.209	<0.001*	0.095
Execution Time (s)	0.188	0.229	0.847	0.890	<0.001*	0.002*

Table 5 P-values from Wilcoxon signed-rank tests comparing FH-FAM to competitors. Asterisks (*) indicate statistical significance ($p < 0.05$). For accuracy and feature reduction, tests assess if FH-FAM is superior (higher values); for execution time, if FH-FAM is faster (lower values).

For classification accuracy, FH-FAM's improvements are statistically significant over FCFB, CFS, DSCA, and FFAMFS ($p < 0.05$), but not over FROT and ABESS. In feature reduction, FH-FAM significantly outperforms CFS, DSCA, and FROT. For execution time, FH-FAM is significantly slower than FROT and ABESS, reflecting the computational cost of fuzzy hypergraph modeling, while differences with other methods are not significant. These tests affirm that FH-FAM's advantages in accuracy and feature reduction are not attributable to chance in key comparisons, strengthening its value for high-dimensional tasks in agriculture and remote sensing.

5.5 Discussion

Table 6 compares FH-FAM with FCFB, CFS, DSCA, FFAMFS, FROT, and ABESS. FH-FAM achieves the highest mean accuracy (81.43%, second-highest rank 7.1, behind only ABESS) and the best mean feature reduction (89.28%), ranking second in Win/Draw/Loss profile thanks to its fuzzy hypergraph framework that captures multi-feature interactions and enables fully automated selection. Execution time is average (11.08 s), balancing efficiency with higher-order modeling complexity. Overall, FH-FAM offers an excellent trade-off in accuracy, reduction, and runtime, making it highly suitable for high-dimensional agriculture and remote sensing tasks. FH-FAM excels when high accuracy and strong feature reduction are needed with reasonable runtime, addressing key literature challenges while outperforming competitors.

Dataset	FH-FAM	FCFB	CFS	DSCA	FFAMFS	FROT	ABESS
Mean accuracy	**81.43**	69.49	66.68	68.33	76.26	81.41	79.24
Mean rank(accuracy)	**7.1**	10	7.33	10	7.2	10	10
Mean reduction	**88.88**	83.31	63.66	56.26	84.34	51.71	84.77
Mean execution time (secs.)	11.80	17.19	9.1	3.69	6.58	**0.31**	0.45

Table 6 Overall comparison of performance.

In application contexts, FH-FAM's strengths in modeling complex, uncertain relationships make it particularly effective for *Food and Agriculture* tasks, where it outperforms recent fuzzy graph methods like FFAMFS by incorporating hypergraphs for better redundancy elimination. For *Remote Sensing*, it addresses high-dimensionality issues more robustly than sparse regression in ABESS, providing a balanced solution for classification in noisy, multi-spectral environments.

6 Conclusion

High-dimensional datasets in agriculture and remote sensing suffer from the curse of dimensionality, complicating classification tasks such as crop variety identification, fruit quality assessment, land cover mapping, and environmental monitoring. Existing feature selection methods often fail to model higher-order interactions or handle inherent uncertainty effectively.

The proposed FH-FAM algorithm addresses these gaps by integrating fuzzy hypergraphs with multi-way mutual information, correlation-based redundancy measures, and staged maximal independent set refinement. Extensive experiments on 15 diverse datasets demonstrate that FH-FAM achieves the highest mean classification accuracy (81.43%) and feature reduction (89.28%) among compared methods (FFAMFS, CFS, DSCA, FCFB, FROT, ABESS), with acceptable execution time. These results highlight FH-FAM's robustness in capturing complex multi-feature dependencies and uncertainty typical in biological/agricultural and multi-spectral remote sensing data.

7 Future work

Future extensions of FH-FAM will address current limitations and expand its utility in high-dimensional domains. These include: supporting non-uniform hyperedge cardinalities (variable-sized hyperedges) to better capture heterogeneous higher-order feature interactions in agricultural and remote sensing datasets [1, 8, 28]; adapting the method for unsupervised and semi-supervised learning to handle limited or unlabeled data, building on existing hypergraph-based techniques in these paradigms; improving scalability for ultra-high-dimensional datasets ($> 10,000$ features) via hyperedge sampling, approximation algorithms, sparse representations, or low-rank hypergraph constructions; exploring incremental/online learning adaptations for streaming data in real-time sensor networks and continuous remote sensing; integrating FH-FAM as a pre-filtering step in deep learning pipelines (e.g., before CNNs or vision transformers for remote sensing image analysis); and optimizing for edge-device deployment to enable low-latency, real-time inference in precision agriculture tasks such as on-the-go crop monitoring and stress detection. These directions will greatly enhance FH-FAM's robustness and applicability to emerging challenges in environmental monitoring, smart farming, and large-scale high-dimensional problems.

Appendix A Derivation of joint-entropy and multi-way mutual information for k-variables

The concept of entropy, introduced by Shannon [29], provides a foundational framework for quantifying uncertainty in probabilistic systems. For two random variables X and Y, the *joint entropy* is given by Eq. A1:

$$H(X,Y) = -\sum_{x}\sum_{y} p(x,y) log[p(x,y)]. \tag{A1}$$

Here, the summation runs over all possible values x of X and y of Y. This 2-variable formula also appears in the foundational work by Shannon.

The *chain rule for entropy* for random variables $X_1, X_2, \cdots, X_k$ is defined in Eq. A2, where each conditional entropy term is defined in Eq. A3.

$$\begin{aligned} H(X_1, X_2, \ldots, X_k) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) \\ + \cdots + H(X_k \mid X_1, X_2, \ldots, X_{k-1}). \end{aligned} \tag{A2}$$

$$H(X_i|X_1, \cdots, X_{i-1}) = -\sum_{x_1,\cdots,x_i} p(x_1, \cdots, x_i) log\ p(x_i|x_1, \cdots, x_{i-1}). \tag{A3}$$

Recalling from the chain rule or factorization rule for joint probabilities Eq. A4 and Eq. A5 can be derived.

$$p(x_1, x_2, \cdots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \cdots p(x_n|x_1, x_2, \cdots, x_{n-1}). \tag{A4}$$

$$p(x_i|x_1, \cdots, x_{i-1}) = \frac{p(x_1, \cdots, x_i)}{p(x_1, \cdots, x_{i-1})}. \tag{A5}$$

This relationship will be crucial when summing over all values to derive the general formula.
Using the chain rule, Eq. A6 can be defined.

$$H(X, Y, Z) = H(X, Y) + H(Z|X, Y). \tag{A6}$$

From the 2-variable formula, Eq. A7 can be expressed.

$$H(X, Y) = -\sum_{x,y} p(x, y)\ log[p(x, y)]. \tag{A7}$$

The conditional entropy term $H(Z|X, Y)$ is defined in Eq. A8.

$$H(Z|X, Y) = -\sum_{x,y,z} p(x, y, z)\ log[p(z|x, y)]. \tag{A8}$$

Since $p(z|x, y) = \frac{p(x,y,z)}{p(x,y)}$, Eq. A9 can be derived.

$$H(Z|X, Y) = -\sum_{x,y,z} p(x, y, z)\ log[\frac{p(x, y, z)}{p(x, y)}]. \tag{A9}$$

Equating the terms from Eq. A7, A8 and A9, Eq. A10 is expressed.

$$\begin{aligned} H(X, Y, Z) &= -\sum_{x,y} p(x, y) \log[p(x, y)] \\ &\quad - \sum_{x,y,z} p(x, y, z) \log\left[\frac{p(x, y, z)}{p(x, y)}\right] \\ &= -\sum_{x,y,z} p(x, y, z) \log[p(x, y, z)]. \end{aligned} \tag{A10}$$

This establishes the 3-variable case.
The mutual information for two random variables X and Y is defined in Eq. A11.

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \tag{A11}$$

where $H(\cdot)$ is the (Shannon) entropy and $H(X, Y)$ is the joint entropy. This appears in Shannon's original paper [29] as well.
With the base case, $k = 2$, Eq. A12 is expressed.

$$I(X_1; X_2) = H(X_1) + H(X_2) - H(X_1, X_2). \tag{A12}$$

And for ($k = 3$), Eq. A13 can be written.

$$I(X_1; X_2; X_3) = H(X_1) + H(X_2) + H(X_3) - H(X_1, X_2, X_3). \tag{A13}$$

Declarations

- Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- Data availability: The datasets analyzed in this study are publicly available from their respective sources as cited in the text. All data were used in accordance with their respective licensing terms, and no additional data were generated in this study.
- Author contribution: Conceptualization, SDC, ARC, AKD; methodology, RD, SB, SDC, ARC, AKD; software, RD, SB; validation, RD, SB, SDC, ARC, AKD; formal analysis, RD, SB; investigation, RD, SB; resources, RD, SB; data curation, RD, SB; writing—original draft preparation, RD, SB, SDC, ARC, AKD; writing—review and editing, RD, SB, SDC, ARC, AKD; visualization, RD, SB; supervision, SDC, ARC, AKD; project administration, SDC, ARC, AKD; funding acquisition, SDC. All authors have read and agreed to the published version of the manuscript.

References

[1] Yuan, M., Yin, C., Hu, J., Zhao, Y., Wang, J., Li, H.: A novel fuzzy-rule-based deep fusion of hypergraph multi-modal for alzheimer's disease detection. Neurocomputing **650**, 130855 (2025)

[2] Baghini, A.Z., Babaei, H., Mirhosseini, R.T.: Determining the membership degree by fuzzy hypergraphs. Mathematical Foundations of Computing **8**(6), 890–904 (2025)

[3] Das, A.K., Goswami, S., Chakrabarti, A., Chakraborty, B.: A new hybrid feature selection approach using feature association map for supervised and unsupervised classification. Expert Systems with Applications **88**, 81–94 (2017)

[4] Das, A.K., Chakraborty, B., Goswami, S., Chakrabarti, A.: A fuzzy set based approach for effective feature selection. Fuzzy Sets and Systems **449**, 187–206 (2022)

[5] Liu, K., Yang, X., Ding, W., Ju, H., Li, T., Wang, J., Yin, T.: A survey on rough feature selection: Recent advances and challenges. IEEE/CAA Journal of Automatica Sinica **6**, 1–22 (2025)

[6] Antelmi, A., Cordasco, G., Polato, M., Scarano, V., Spagnuolo, C., Yang, D.: A survey on hypergraph representation learning. ACM Computing Surveys **56**(1), 1–38 (2023)

[7] Feng, D., Liu, W., Peng, S., Liu, Y., Zhou, W.: Corofr: Community detection of feature graph improves feature selection using robust fuzzy rough set. Advanced Intelligent Systems, 2500044 (2025)

[8] Yin, T., Ding, W., Ju, H., Huang, J., Chen, Y.: The fuzzy hypergraph neural network model based on sparse k-nearest neighborhood granule. Applied Soft Computing **170**, 112721 (2025)

[9] Koklu, M., Kursun, R., Taspinar, Y.S., Cinar, I.: Classification of date fruits into genetic varieties using image analysis. Mathematical Problems in Engineering **2021**(1), 4793293 (2021)

[10] Koklu, M., Ozkan, I.A.: Multiclass classification of dry beans using computer vision and machine learning techniques. Computers and Electronics in Agriculture **174**, 105507 (2020)

[11] Kaya, E., Saritas, I.: Towards a real-time sorting system: Identification of vitreous durum wheat kernels using ann based on their morphological, colour, wavelet and gaborlet features. Computers and Electronics in Agriculture **166**, 105016 (2019)

[12] Singh, D., Taspinar, Y.S., Kursun, R., Cinar, I., Koklu, M., Ozkan, I.A., Lee, H.-N.: Classification and analysis of pistachio species with pre-trained deep learning models. Electronics **11**(7), 981 (2022)

[13] Koklu, M., Sarigil, S., Ozbek, O.: The use of machine learning methods in classification of pumpkin seeds (cucurbita pepo l.). Genetic Resources and Crop Evolution **68**(7), 2713–2726 (2021)

[14] Çınar, İ., Koklu, M., Taşdemir, Ş.: Classification of raisin grains using machine vision and artificial intelligence methods. Gazi Journal of Engineering Sciences **6**(3), 200–209 (2020)

[15] Cinar, I., Koklu, M.: Identification of rice varieties using machine learning algorithms. Journal of Agricultural Sciences, 9–9 (2022)

[16] Blackard, J.: Covertype. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C50K5N (1998)

[17] Crop mapping using fused optical-radar data set. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5G89D (2020)

[18] EUROCROP24: Crop Health and Environmental Stress Dataset. Kaggle (2025). https://www.kaggle.com/dsv/10605890

[19] Johnson, B.: Forest type mapping. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5QP56 (2012)

[20] Srinivasan, A.: Statlog (Landsat Satellite). UCI Machine Learning Repository.

DOI: https://doi.org/10.24432/C55887 (1993)

[21] Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P.: Global land use/land cover with sentinel 2 and deep learning. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 4704–4707 (2021). IEEE

[22] Narayan, N.: Weather Type Classification. Online. Accessed: 2025-09-02 (2024). https://www.kaggle.com/datasets/nikhil7280/weather-type-classification

[23] Hall, M.A.: Correlation-based feature selection for machine learning. PhD thesis, The University of Waikato (1999)

[24] Zhang, Z., Hancock, E.R.: A graph-based approach to feature selection. In: International Workshop on Graph-based Representations in Pattern Recognition, pp. 205–214 (2011). Springer

[25] Yu, L., Liu, H.: Feature selection for high-dimensional data: A fast correlation-based filter solution. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 856–863 (2003)

[26] Petrovich, M., Liang, C., Sato, R., Liu, Y., Tsai, Y.-H.H., Zhu, L., Yang, Y., Salakhutdinov, R., Yamada, M.: Feature-robust optimal transport for high-dimensional data. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 291–307 (2022). Springer

[27] Zhu, J., Wang, X., Hu, L., Huang, J., Jiang, K., Zhang, Y., Lin, S., Zhu, J.: abess: A fast best-subset selection library in python and r. Journal of Machine Learning Research **23**(202), 1–7 (2022)

[28] Zheng, Z., Geng, S., Zhou, M.: Research on feature selection methods based on hypergraph theory in multi-scale decision information system. In: Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Intelligent Information Processing, pp. 258–264 (2024)

[29] Shannon, C.E.: A mathematical theory of communication. The Bell system technical journal **27**(3), 379–423 (1948)