AI

MDPI

Article

A Fuzzy and Explainable AI Framework for Comparing Physical and Perceptual Representations in Galaxy Morphology

Gabriel Marín Díaz [1,2,*], Alvaro Manuel Rodriguez-Rodriguez [2] and Eva María Andrés Núñez [2]

[1] Faculty of Statistics, Complutense University, Puerta de Hierro, 28040 Madrid, Spain
[2] Science and Aerospace Department, Universidad Europea de Madrid, 28670 Villaviciosa de Odón, Madrid, Spain; alvaromanuel.rodriguez@universidadeuropea.es (A.M.R.-R.); evamaria.andres@universidadeuropea.es (E.M.A.N.)
* Correspondence: gmarin03@ucm.es

Abstract

Galaxy morphology combines measurable structural properties with subjective visual interpretation, limiting strictly hard-label classifications. This study proposes a framework designed to compare physically derived and human-based galaxy classifications while explicitly accounting for uncertainty and interpretability. Using photometric and structural features from the Sloan Digital Sky Survey (SDSS), physical groupings are obtained through Fuzzy C-Means clustering, enabling gradual transitions via soft memberships. Human clusters are constructed from Galaxy Zoo 2 debiased vote fractions, capturing aggregated perceptual judgments. Supervised models are trained to predict both physical and human cluster assignments from the same set of physical variables, providing a quantitative assessment of structural coherence and perceptual–physical alignment. SHAP-based explainability identifies the relative influence of color and concentration parameters in each scheme. Results show that physical clustering is driven by structural concentration and bulge dominance, while human classification exhibits smoother decision boundaries and greater sensitivity to photometric appearance. Discrepancies concentrate in transitional and orientation-sensitive systems. An interactive visualization layer supports traceable qualitative inspection. The framework provides a reproducible methodology for analyzing classification consistency, uncertainty, and human–model alignment.

Keywords: Fuzzy C-Means; explainable artificial intelligence; galaxy morphology; photometric features; human perceptual classification; SDSS

MSC: 62H30; 85A35

Academic Editor: Tihomir Orehovački

Received: 20 March 2026
Revised: 18 April 2026
Accepted: 20 April 2026
Published: 30 April 2026

1. Introduction

Galaxy morphology has long been studied from two complementary perspectives: quantitative structural measurements derived from large observational surveys and visual classifications provided through human inspection. Surveys such as the Sloan Digital Sky Survey (SDSS) offer detailed photometric and structural descriptors while citizen science initiatives such as Galaxy Zoo demonstrate that collective human perception can capture morphological subtleties not always reflected in global physical parameters [1,2]. These two approaches, however, are not equivalent. Physical measurements are often interpreted through discrete partitions imposed on intrinsically continuous structures, whereas human classifications are inherently probabilistic and subject to perceptual uncertainty.

Recent developments in artificial intelligence have emphasized the importance of explicitly modeling uncertainty and interpretability, particularly in classification problems where boundaries are gradual rather than sharply defined [3]. In this context, fuzzy clustering provides a natural mechanism to represent transitional regimes through soft memberships, while explainable AI techniques enable transparent analysis of the contribution of individual features. In the present study, uncertainty is understood primarily in two senses: as membership uncertainty in fuzzy clustering, where objects may belong partially to multiple regimes, and as perceptual uncertainty in human voting, where Galaxy Zoo classifications are represented by probabilistic vote fractions rather than fixed labels.

Despite the availability of both physical and perceptual classifications, systematic comparisons between these representations remain limited [4,5]. It is still unclear to what extent structural photometric properties align with, approximate, or diverge from collective human morphological judgments, and how such differences should be interpreted.

In this work, we propose an interpretable framework to analyze the relationship between physical measurements and human perception in galaxy morphology. By combining fuzzy clustering with supervised modeling and SHAP-based feature attribution, we quantify both the degree of agreement between structural and perceptual classifications and the variables that drive their correspondence or divergence. An interactive visualization component linking clustering outcomes to SDSS imagery supports object-level inspection and facilitates human interpretation of ambiguous and transitional cases.

The objective of this study is not to define a definitive morphological taxonomy, but rather to provide a methodological framework for examining consistency, uncertainty, and transitional regimes between physically derived and human-based classification systems.

The remainder of this paper is organized as follows. Section 2 reviews related work on physical and perceptual classification in astronomy and recent advances in fuzzy clustering and explainable AI. Section 3 describes the proposed methodology. Section 4 presents the results and visual validation. Section 5 discusses the findings and outlines future directions. Section 6 concludes the study.

2. Related Work

2.1. Parallel Evolution of Three Methodological Traditions

Over the past decade, three research directions have progressed largely in parallel: fuzzy clustering methodologies, explainable artificial intelligence (XAI), and investigations into discrepancies between human perceptual classification and measurable physical properties.

Fuzzy clustering methods, particularly Fuzzy C-Means (FCM) [6], were originally conceived to model gradual transitions and overlapping class memberships. Unlike hard clustering approaches, FCM assigns degrees of membership, making it conceptually suitable for systems characterized by continuity, ambiguity, or intrinsic uncertainty, properties commonly observed in complex physical domains where boundaries are not sharply defined.

In parallel, the widespread adoption of deep learning models raised increasing concerns regarding interpretability, leading to the emergence of explainable artificial intelligence as a structured response to the opacity of high-capacity predictive systems [7]. While interpretability has become central in many domains, its integration with uncertainty-aware methodologies remains relatively limited. A Web of Science topic search combining fuzzy clustering with interpretability-related terms reveals a modest but noticeable increase in publications after 2020 (68 records, 376 citations), suggesting a growing but still emerging convergence between these areas (Figure 1).

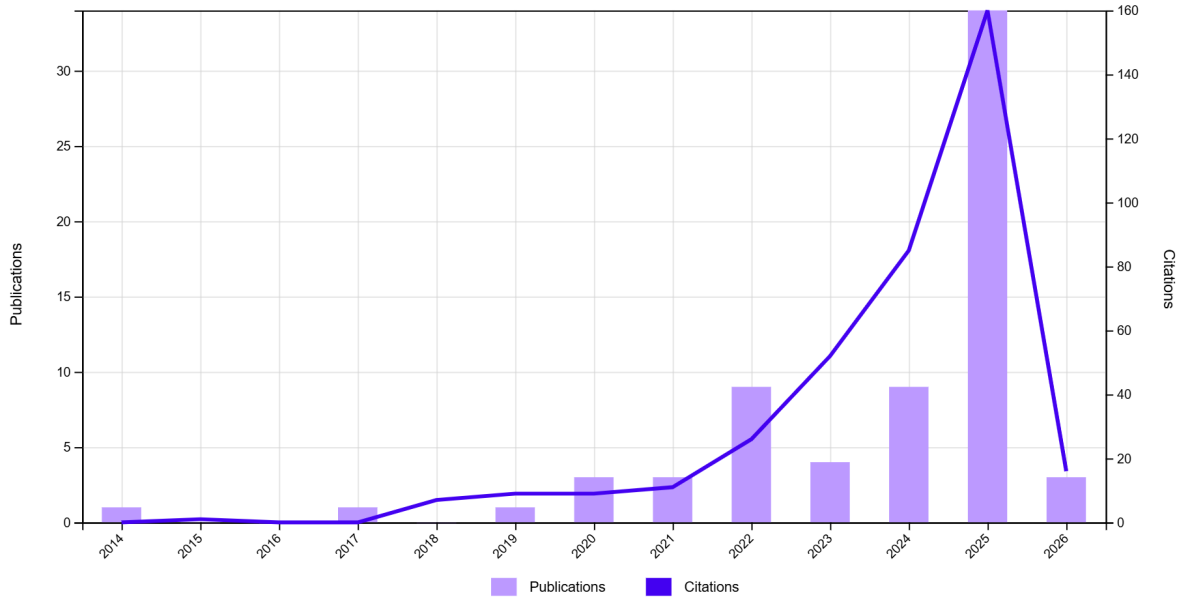

Figure 1. Publications (68) and Citations (376). TS = ((“explainable artificial intelligence” OR “XAI” OR “interpretability”) AND (“Fuzzy Clustering” OR “Fuzzy C-Means”)).

A third line of research has examined the relationship between human classification and quantitative physical measurements. A search combining human perception or classification with physical properties retrieves a comparatively small body of literature (18 publications, 301 citations; Figure 2). These studies consistently report divergences between perceptual judgments and measurable descriptors; however, they rarely incorporate structured uncertainty modeling or interpretable machine learning as analytical tools.

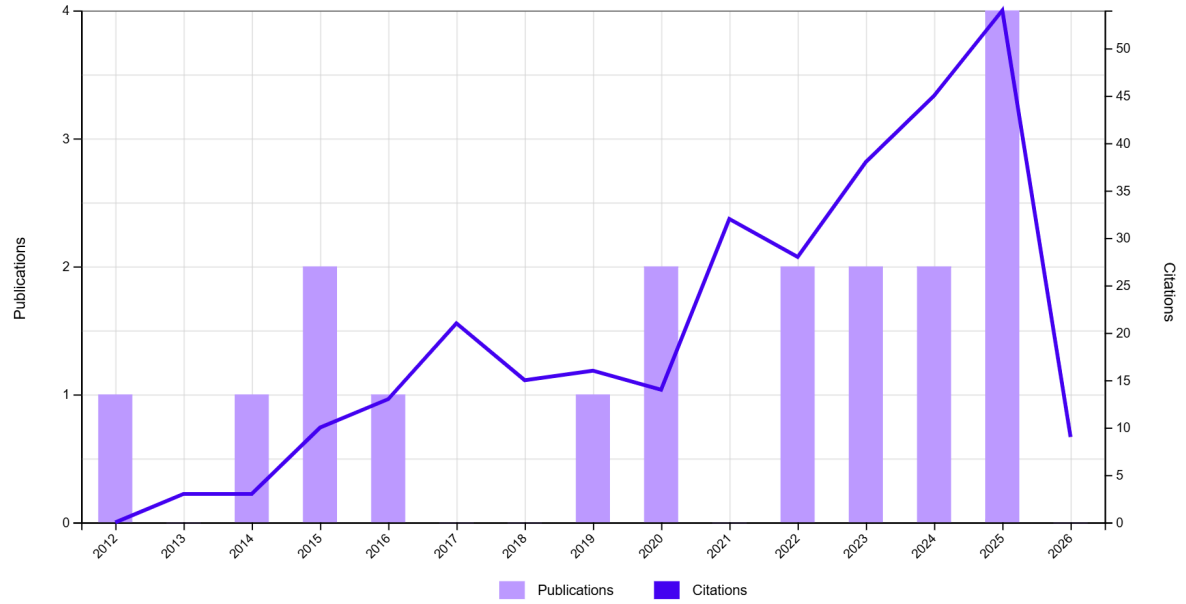

Figure 2. Publications (18) and Citations (301). TS = ((“human classification” OR “human perception”) AND (“physical properties” OR “physical parameters”)).

Although conceptually complementary, these three traditions have largely evolved independently, with limited efforts toward systematic methodological integration.

2.2. Machine Learning in Astronomy: Performance over Interpretability

In astronomy and cosmology, machine learning has experienced substantial expansion over the last decade. Convolutional neural networks, in particular, have been widely adopted for galaxy morphology classification, frequently leveraging large citizen science datasets such as Galaxy Zoo [1,2,8].

Beyond morphology, deep learning has been applied to photometric redshift estimation, anomaly detection in wide-field surveys, cosmological parameter inference from large-scale structure simulations, and reconstruction and analysis of Cosmic Microwave Background (CMB) maps [9–11]. These approaches demonstrate strong predictive capabilities and scalability across massive datasets.

Nevertheless, most implementations rely on highly complex architectures whose internal representations are difficult to interpret. When explainability techniques are introduced, they are typically applied post hoc to supervised models rather than incorporated into inherently interpretable or uncertainty-aware analytical frameworks.

In contrast, fuzzy methodologies remain sparsely represented in astronomical research. Restricting bibliometric searches to fuzzy clustering within astronomical contexts retrieves only four publications (84 citations; Figure 3), distributed over time without evidence of sustained growth [12–16]. This limited adoption is noteworthy, given that astronomical systems frequently exhibit gradual morphological transitions, overlapping structural classes, and significant observational uncertainties, conditions under which fuzzy approaches are theoretically well suited.

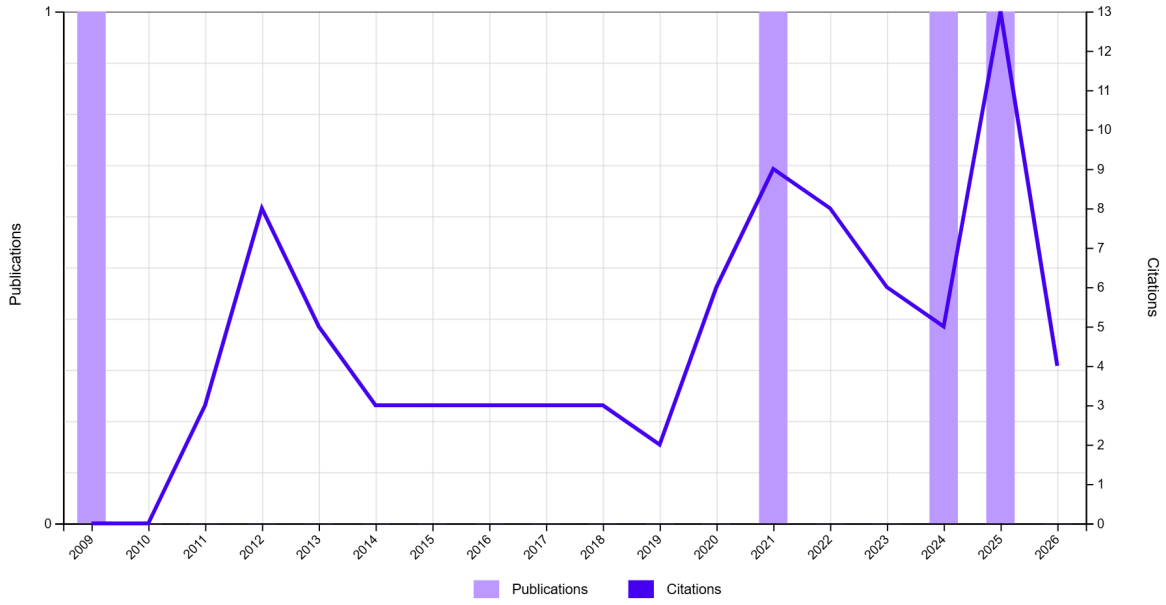

Figure 3. Publications (4) and Citations (84). TS = ((“Fuzzy C-Means” OR “fuzzy clustering”) AND (“astronomy” OR “astrophysics” OR “cosmology” OR “galaxy classification”)).

Recent interest in XAI within astronomy shows clearer growth, with 31 publications and 345 citations identified through topic searches (Figure 4). However, these contributions predominantly focus on interpreting outputs from supervised deep learning models, and their integration with uncertainty-aware methodologies remains largely unexplored.

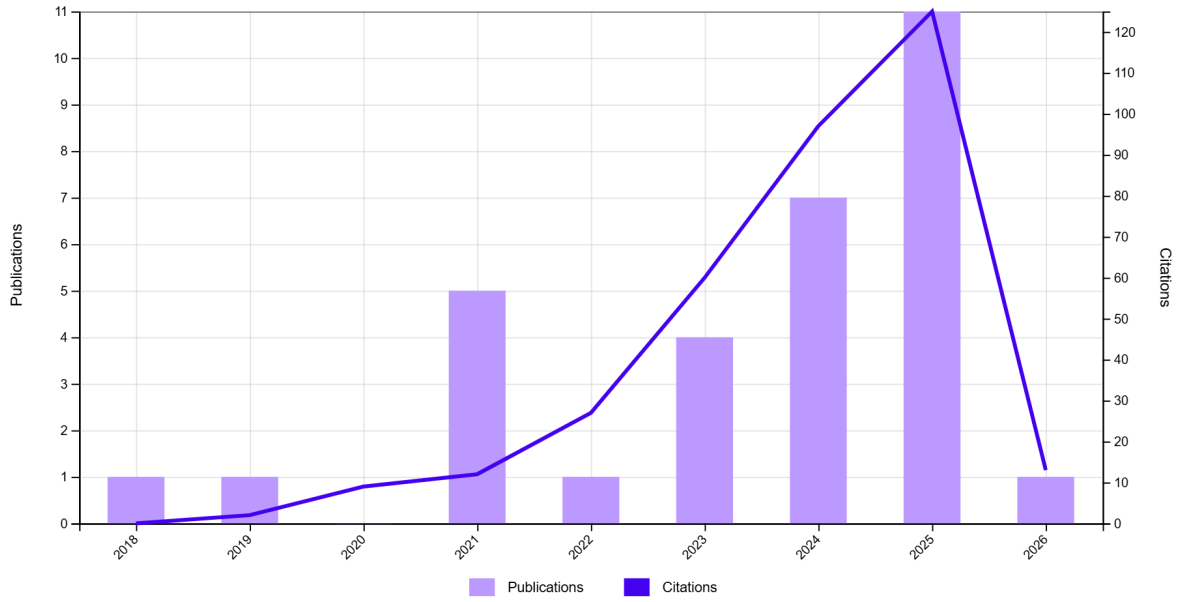

Figure 4. Publications (31) and Citations (345). TS = ((“explainable artificial intelligence” OR “XAI” OR “interpretability”) AND (“astronomy” OR “astrophysics” OR “cosmology” OR “galaxy classification”)).

2.3. Toward an Integrated Uncertainty–Interpretability Framework

When fuzzy clustering, explainable AI, and astronomical applications are combined within a single bibliometric query, only two records are retrieved, highlighting the limited development of this intersection. In ref. [13], fuzzy clustering and explainable AI are integrated in an astronomical context, but the study does not explicitly address the relationship between physically derived galaxy groupings and aggregated human perceptual classifications within a unified analytical framework. The work therefore reflects an initial step toward methodological convergence, rather than a structured and generalizable framework for combining fuzzy uncertainty modeling with explainable machine learning in astrophysical contexts.

Similarly, incorporating machine learning into studies of perceptual–physical discrepancies yields only three publications [17–19], none of which address astronomical morphology or combine fuzzy logic with XAI methodologies. These studies are mainly situated in other applied domains and focus on perception-aware modeling or soft metrology, rather than on uncertainty-aware comparisons between human visual judgments and physically measurable structure in galaxy classification.

Taken together, these findings point to a clear methodological gap. Although fuzzy clustering, explainable AI, deep learning in astronomy, and perceptual–physical comparison studies have each advanced in parallel, there is still no unified framework that reconciles human morphological classifications with physically derived parameters through an interpretable, uncertainty-aware analytical structure. More specifically, previous studies have tended to address only part of the problem: some incorporate uncertainty but not interpretability, others include interpretability but not explicit fuzzy modeling, and others examine the relationship between perception and physical properties without providing a joint framework for systematically analyzing agreement, disagreement, and transitional cases.

This gap is particularly relevant in astronomical morphology, where classifications often depend on human perception, frequently through citizen-science labeling, while physical parameters such as redshift, stellar mass, color indices, and structural descriptors provide complementary quantitative information. Understanding how these representa-

tions align, diverge, and transition across ambiguous regimes requires a framework that jointly addresses uncertainty, interpretability, and the relationship between perceptual and physical classifications.

Addressing this need motivates integrative methodologies such as the Fuzzy-Adaptive System for Explainable AI (FAS-XAI) framework [20], which enables systematic analysis of agreement, discrepancy, and transitional structures between perceptual and physically derived classifications.

3. Materials and Methods

This section describes the methodological design adopted to comparatively analyze physically derived galaxy properties and human morphological classifications. The objective is to establish a structured analytical procedure that integrates fuzzy clustering, supervised modeling, and explainable artificial intelligence techniques to examine the relationship between structural measurements and perceptual labeling.

Rather than treating physical and human classifications independently, the proposed approach analyzes both domains within a unified uncertainty-aware framework. Physical parameters derived from spectroscopic and photometric measurements are evaluated alongside aggregated citizen science morphological annotations. By modeling each representation separately and subsequently comparing their structures, the methodology enables systematic assessment of convergence, divergence, and transitional regimes between perceptual and physically grounded descriptions.

The workflow is organized into a sequence of interconnected phases, each addressing a specific analytical objective, from data harmonization to interpretability assessment and visual validation. Figure 5 summarizes the overall methodological architecture.

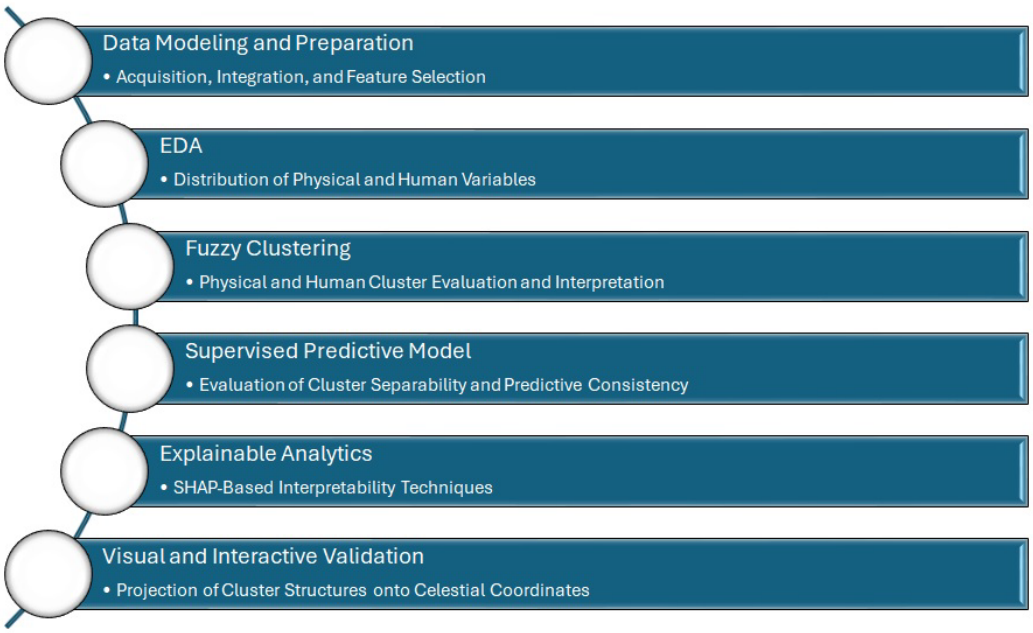

Figure 5. Methodological framework.

The methodological process unfolds as follows:

- **Data Modeling and Preparation.** Physical measurements obtained from SDSS are combined with human morphological annotations from Galaxy Zoo 2 using unique object identifiers to ensure one-to-one correspondence. Representative variables from both domains are selected based on relevance and completeness. All features are normalized to ensure scale comparability, and extreme outliers are removed to re-

duce distortion in subsequent analyses. This stage establishes a consistent basis for comparative modeling.

- **Exploratory Data Analysis (EDA).** Prior to clustering, both physical and perceptual variables are examined through descriptive statistics and distributional analysis. Histogram-based inspection and dispersion measures are used to identify skewness, multimodality, and structural heterogeneity. This step informs clustering parameter selection and provides contextual understanding of variable behavior.
- **Fuzzy Clustering.** Fuzzy C-Means (FCM) clustering is applied separately to the physical feature space and the human annotation space. By assigning graded membership values rather than hard labels, FCM captures transitional regimes between morphological categories. Cluster centroids and membership distributions are analyzed to characterize dominant patterns. A correspondence analysis is then performed to evaluate alignment between physical clusters and perceptual groupings.
- **Supervised Predictive Model.** XGBoost classifiers are trained to assess cluster separability and predictive consistency using identical physical inputs. This stage quantifies the internal coherence of physical clustering and evaluates how well structural variables reproduce human morphological classifications.
- **Explainable Analytics.** Model interpretability is examined using SHAP-based feature attribution. Both global importance rankings and local contribution analyses are computed. This step identifies the physical variables that most strongly influence cluster assignments and morphological predictions, clarifying the structural drivers underlying each representation.
- **Visual and Interactive Validation.** Cluster assignments and membership degrees are projected onto celestial coordinates (RA–DEC) to inspect spatial distributions. An interactive visualization interface enables object-level exploration, linking SDSS imagery with physical attributes and fuzzy memberships. Representative galaxy images are also analyzed for each cluster correspondence category (1–1, 1–0, 0–1, 0–0), where the first index denotes physical clustering and the second corresponds to human-based classification. This supports qualitative validation and facilitates interpretation of transitional or ambiguous cases.

Through this workflow, uncertainty modeling, comparative analysis, predictive validation, and interpretability assessment are integrated within a coherent analytical framework. The procedure enables systematic examination of correspondences and discrepancies between quantitative physical measurements and human morphological perception.

3.1. Data Modeling and Preparation

This phase establishes the structural foundation of the proposed analytical framework. Rather than relying on predefined supervised morphology labels, the study integrates two complementary representations derived from the Galaxy Zoo project.

The analysis is based on the publicly available Galaxy Zoo 2 (GZ2) data release, which provides detailed morphological classifications and associated Sloan Digital Sky Survey (SDSS) photometric and structural parameters for nearly 300,000 galaxies. The GZ2 project description, data reduction procedures, and debiasing methodology are presented in Willett et al. (2013) [21].

The physical measurements employed in this work, including redshift, photometric color indices, concentration index, and fractional de Vaucouleurs component, originate from SDSS observations and are distributed within the GZ2 catalog. These variables provide quantitative descriptors of stellar populations and structural morphology. Together, they define a compact and interpretable set of physical features designed for direct comparison with the human perceptual representation.

Human perceptual information is represented through debiased vote fractions derived from the GZ2 hierarchical classification tree. These probabilities encode the aggregated judgments of volunteer classifiers across multiple morphological dimensions, including smoothness, spiral structure, bar presence, bulge prominence, and shape characteristics.

The objective of this phase is to construct two well-defined feature spaces: a physical feature space based on measurable photometric and structural parameters, and a perceptual feature space based on aggregated human morphological probabilities. These spaces are modeled independently and subsequently compared within a unified fuzzy analytical framework, enabling systematic evaluation of correspondence and divergence between structural measurements and collective visual interpretation.

3.1.1. Physical Feature Space Construction

Each galaxy i is represented by a vector of physically meaningful attributes extracted from SDSS and distributed through the GZ2 catalog.

- **Redshift (z_i)** measures the cosmological displacement of spectral lines and provides evolutionary context for the observed galaxy structure.
- **Color index ($(g-r)_i$)** is defined as the difference between magnitudes in the SDSS g (green) and r (red) photometric bands. Redder galaxies typically host older stellar populations, while bluer galaxies indicate ongoing star formation.
- **Color index ($(u-r)_i$)** measures the magnitude difference between the ultraviolet u band and the r band, providing sensitivity to recent star formation and facilitating separation between early, and late, type systems.
- **Concentration index ($C_{r,i}$)** is defined as follows:

$$C_{r,i} = \frac{R_{90}}{R_{50}} \tag{1}$$

where R_{90} and R_{50} denote the radii enclosing 90% and 50% of the Petrosian flux in the r band. Larger values indicate centrally concentrated light distributions.

- **Fractional de Vaucouleurs component ($fracDeV_{r,i}$)** represents the fraction of r-band light best fitted by a de Vaucouleurs profile rather than an exponential disk profile.

The physical representation of galaxy i is therefore as follows:

$$x_i^{(phys)} = \left[z_i,\ (g-r)_i,\ (u-r)_i,\ C_{r,i},\ fracDeV_{r,i} \right] \in \mathbb{R}^5 \tag{2}$$

The complete physical dataset is as follows:

$$X^{(phys)} \in \mathbb{R}^{N \times 5}, \tag{3}$$

where N denotes the number of galaxies retained after preprocessing.

These variables were selected due to their established relationship with galaxy morphology, jointly capturing stellar population properties, structural concentration, bulge dominance, and evolutionary stage.

3.1.2. Human Perceptual Feature Space Construction

Human morphological perception is encoded using debiased vote fractions from the Galaxy Zoo 2 decision tree. For each galaxy i,

$$p_{ij} \in [0, 1]$$

denotes the debiased fraction of classifiers selecting perceptual response j.

Twelve representative perceptual variables were selected, including smooth morphology probability, edge-on probability, bar presence probability, spiral structure probability, bulge prominence levels, shape descriptors (completely round, in-between, cigar-shaped), and ring probability. Together, they define a compact and interpretable perceptual space that captures the main visual dimensions of galaxy morphology in Galaxy Zoo 2. The aim was not to reproduce the complete annotation tree, but to define a perceptual space directly comparable with the physical representation.

The perceptual representation is therefore as follows:

$$x_i^{(human)} = [p_{i1}, p_{i2}, \ldots, p_{iM}] \in \mathbb{R}^M, \text{where } M = 12 \quad (4)$$

The full perceptual dataset is defined as follows:

$$X^{(human)} \in \mathbb{R}^{N\ x\ 12} \quad (5)$$

The probabilistic nature of these variables inherently reflects classification uncertainty, which is compatible with fuzzy modeling.

3.1.3. Dataset Harmonization and Preprocessing

The physical and perceptual datasets were merged using a shared object identifier (OBJID–dr7objid). An inner join operation ensured that only galaxies with complete physical measurements and human annotations were retained. The resulting harmonized dataset contains 243,500 galaxies. Preprocessing steps included:

- Removal of incomplete records.
- Filtering of extreme values to reduce distortion in distance-based clustering.
- Standardization of selected variables prior to fuzzy modeling.

Since Fuzzy C-Means relies on Euclidean distance, all features were standardized using z-score normalization:

$$x_{ij}^{norm} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad (6)$$

where μ_j and σ_j denote the mean and standard deviation of feature j, respectively.

This normalization prevents scale imbalances from dominating the clustering objective function.

At the conclusion of this phase, the dataset consists of:

- Five normalized physical variables per galaxy (plus RA and DEC for visualization).
- Twelve normalized perceptual variables per galaxy.
- A unique object identifier enabling traceability.

This prepared dataset serves as the input for the fuzzy clustering procedures described in the following section.

3.2. Exploratory Data Analysis (EDA)

Prior to normalization and fuzzy modeling, an exploratory data analysis (EDA) was conducted to examine the statistical structure of both physical and perceptual variables in their original measurement scales. This stage aims to characterize distributional behavior, identify dispersion patterns and skewness, detect potential multimodality, and evaluate inter-variable relationships that may influence subsequent modeling. Such inspection is particularly relevant, as fuzzy clustering relies on distance-based metrics that can be sensitive to asymmetries and feature correlations.

The EDA was performed independently for the physical and perceptual feature spaces and comprised two complementary analyses.

First, marginal distributions were examined through histogram-based visualization. This step enables identification of skewed distributions, concentration effects, heavy tails, and polarization patterns that may influence normalization and cluster formation.

Second, correlation matrices were computed separately for each feature space using Pearson correlation coefficients:

$$\rho_{jk} = \frac{Cov(X_j, X_k)}{\sigma_j \sigma_k}, \tag{7}$$

where X_j and X_k denote two variables and σ_j, σ_k their respective standard deviations. This analysis provides insight into redundancy, linear dependence, and structural coherence within each domain, informing both the fuzzy clustering stage and subsequent supervised modeling.

3.3. Fuzzy Clustering Process

3.3.1. Fuzzy Clustering

To model gradual morphological transitions and overlapping structural regimes, Fuzzy C-Means (FCM) clustering was independently applied to both the physical and human perceptual feature spaces. Unlike hard clustering approaches, which assign each galaxy to a single class, FCM allows soft memberships, enabling a galaxy to belong to multiple clusters with varying degrees of association [6].

Given a dataset $X = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$, the FCM algorithm partitions the data into c clusters by minimizing the objective function:

$$J_m = \sum_{i=1}^{N} \sum_{k=1}^{c} u_{ik}^m \| x_i - c_k \|^2 \tag{8}$$

where

- $u_{ik} \in [0, 1]$ denotes the membership degree of galaxy i to cluster k;
- $c_k \in \mathbb{R}^d$ is the centroid of cluster k;
- $m > 1$ is the fuzzification parameter;
- $\| \cdot \|$ represents the Euclidean norm.

In this study, the fuzzification parameter was set to $m = 2$, following the standard choice in Fuzzy C-Means applications. This value provides a balanced compromise between overly crisp and overly diffuse memberships, which is appropriate for representing gradual morphological transitions while preserving cluster separability. Euclidean distance was adopted because all variables were standardized prior to clustering, making their scales directly comparable and consistent with the centroid-based formulation of FCM. Under these conditions, Euclidean distance provides a simple and interpretable measure of proximity in the selected feature spaces. At the same time, both choices involve limitations: the value of m influences the sharpness of the membership distribution, and Euclidean distance may be less suitable for correlated variables or non-compact structures. Therefore, the resulting clusters should be interpreted as an uncertainty-aware representation of the dominant patterns in the standardized feature space.

The membership matrix $U = [u_{ik}]$ satisfies the constraint:

$$\sum_{k=1}^{c} u_{ik} = 1, \quad \forall i \in 1, \ldots, N \tag{9}$$

Optimization proceeds iteratively by alternating between centroid updates:

$$c_k = \frac{\sum_{i=1}^{N} u_{ik}^m x_i}{\sum_{i=1}^{N} u_{ik}^m} \tag{10}$$

and membership updates:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}}} \tag{11}$$

until convergence of the objective function is achieved.

For interpretative purposes and correspondence analysis, dominant hard labels were derived as

$$\hat{y}_i = \underset{k}{argmax}\, u_{ik}, \tag{12}$$

while retaining the full fuzzy membership matrix U for uncertainty and overlap analysis. This dual representation preserves soft-clustering information while ensuring compatibility with supervised modeling and cross-tabulation procedures.

3.3.2. Selection of the Number of Clusters

The optimal number of clusters c was evaluated using internal validity indices. The Fuzzy Partition Coefficient (FPC) [22]:

$$\mathrm{FPC} = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{c} u_{ik}^2 \tag{13}$$

where larger values indicate clearer partition structure.

The Xie–Beni index (XB) [23]:

$$\mathrm{XB} = \frac{\sum_{i=1}^{N} \sum_{k=1}^{c} u_{ik}^{m} \| x_i - c_k \|^2}{N \cdot \min_{k \neq j} \| c_k - c_j \|^2} \tag{14}$$

where lower values indicate compact and well-separated clusters.

According to these criteria, $c = 2$ was retained as the most consistent and interpretable configuration for the purposes of comparative analysis. This solution captures a dominant structural dichotomy broadly consistent with early type (bulge-dominated) versus late-type (disk-dominated) morphological regimes while preserving transitional memberships.

3.4. Supervised Predictive Model

To quantitatively evaluate the structural coherence and separability of the fuzzy clusters identified in Section 3.3, supervised classification models based on XGBoost were trained using exclusively physical input features [24]. The objective of this stage is twofold: to assess the internal consistency of the clustering structures and to quantify the correspondence between physically derived and perceptually derived regimes.

Two predictive scenarios were examined. First, dominant hard labels obtained from fuzzy clustering in the physical feature space were predicted using the same physical variables. Second, dominant labels derived from clustering in the human perceptual space were predicted using identical physical inputs.

Let $x_i \in \mathbb{R}^5$ denote the physical feature vector of galaxy i, and $y_i \in \{0, 1\}$ represent the dominant cluster assignment obtained as defined in Equation (12). The supervised model estimates a mapping,

$$f : \mathbb{R}^5 \rightarrow \{0, 1\} \tag{15}$$

optimized through gradient boosting by minimizing a regularized logistic loss function:

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} \ell(y_i, f(x_i)) + \sum_{k=1}^{K} \Omega(T_k), \tag{16}$$

where ℓ denotes the binary cross-entropy loss, T_k represents individual decision trees, Ω is a regularization term controlling model complexity, and K is the number of boosting iterations. Stratified train–test splits were employed to preserve class balance and ensure robust evaluation.

Predictive performance serves distinct interpretative roles. When predicting physical cluster labels, high accuracy indicates that the fuzzy partition is structurally coherent and well separated within the physical feature space. When predicting human-derived cluster labels from physical variables, performance quantifies the extent to which measurable structural properties approximate perceptual morphological classifications. Lower predictive accuracy in this setting suggests the presence of perceptual, nonlinear, or orientation-dependent cues not fully captured by the selected physical descriptors.

This supervised stage provides a quantitative assessment of cluster separability, structural consistency, and cross-domain alignment, establishing the basis for the interpretability analysis presented in the subsequent section.

3.5. Explainable Analytics

To ensure interpretability and transparency of the clustering structures, SHAP (SHapley Additive Explanations) was applied to the supervised XGBoost models introduced in Section 3.4 [25]. Since fuzzy clustering does not inherently provide feature attribution, interpretability is obtained indirectly by analyzing classifiers trained to predict both physical and human cluster assignments from physical input variables.

Let $f(x_i)$ denote the trained supervised model and $x_i \in \mathbb{R}^5$ the physical feature vector of galaxy i. SHAP assigns to each feature j a contribution value $\phi_j^{(i)}$ such that the model output can be expressed as follows:

$$f(x_i) = \phi_0 + \sum_{j=1}^{5} \phi_j^{(i)}, \tag{17}$$

where ϕ_0 represents the expected model output over the dataset and $\phi_j^{(i)}$ measures the marginal contribution of feature j for instance i.

These contributions are computed as Shapley values from cooperative game theory:

$$\phi_j^{(i)} = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{j\}}(x_i) - f_S(x_i) \right], \tag{18}$$

where F denotes the set of all features and S a subset of features excluding j.

3.5.1. Global Interpretability

Global SHAP values were aggregated across all galaxies to evaluate the overall importance of each physical parameter in distinguishing both physical fuzzy clusters and human-derived fuzzy clusters [26]. This analysis identifies which structural variables most strongly contribute to cluster separability and whether similar physical drivers govern both structural and perceptual regimes.

For comparison, model-based importance metrics from XGBoost were also examined, allowing contrast between gain-based importance and SHAP-based attribution [27].

3.5.2. Local Interpretability

At the instance level, local SHAP explanations were computed for selected galaxies, enabling detailed inspection of how individual physical attributes contribute positively or negatively to predicted cluster assignments [28].

Local analysis was particularly informative for:

- Transitional galaxies with balanced fuzzy memberships.
- Cases exhibiting disagreement between physical and human clustering.
- Representative galaxies selected for visual validation.

Instance-level attribution provides a fine-grained understanding of the structural drivers underlying each classification outcome and facilitates interpretation of discrepancies between physically derived and perceptual classifications.

3.6. Visual and Interactive Validation

To complement the quantitative analyses, a visual validation stage was implemented to link clustering outcomes with direct observational evidence. This phase comprises two components: image-based comparison across agreement and disagreement configurations, and an interactive visualization interface enabling object-level inspection in celestial coordinates.

In the first component, galaxies are grouped according to the correspondence between dominant physical and human cluster assignments, producing four cases, (1,1), (1,0), (0,1), and (0,0), where the first index denotes the physical cluster label and the second the human label. For each group, representative SDSS cutout images are displayed to allow qualitative inspection of morphological consistency within agreement regimes ((1,1), (0,0)) and targeted examination of mismatch configurations ((1,0), (0,1)). This comparison facilitates identification of transitional morphologies and systematic divergences between perceptual labeling and physically derived structure.

The second component consists of an interactive dashboard designed for exploratory, human-centered analysis at the individual-galaxy level. Galaxies are projected onto celestial coordinates (RA, DEC) in a Plotly based scatter visualization, with points colored according to either physical or human cluster assignment. Selecting a galaxy retrieves an SDSS DR17 cutout image via the SkyServer ImgCutout service and displays it alongside its associated attributes [29]. The information panel includes the object identifiers (specobjid, OBJID), sky coordinates, physical variables (z, $g-r$, $u-r$, C_r, $fracDeV_r$), and the debiased Galaxy Zoo 2 vote fractions used as perceptual descriptors. A direct SkyServer navigation link is provided to enable external verification.

Together, these visual components support qualitative validation of clustering correspondence and enable traceable inspection of how physical measurements and human annotations relate to fuzzy membership structures. In particular, the interactive analysis environment, implemented in Python using Plotly based visualization tools, facilitates the exploration of ambiguous and transitional cases, supporting human interpretation and enhancing transparency in the analysis.

4. Results

This section presents the empirical results obtained following the methodological procedure described in Section 3. The analysis is organized according to the main stages of the framework, enabling direct comparison between structural measurements and human morphological classifications.

Quantitative metrics and visual evidence are reported at each stage to assess dataset properties, clustering behavior in both feature spaces, cross-domain correspondence, predictive separability, and feature attribution patterns.

Together, these results evaluate the internal consistency of the fuzzy partitions and the extent to which structural parameters physically derived align with human morphological classifications.

4.1. Dataset Overview After Cleaning

Following the integration and preprocessing procedures described in Section 3.1, a harmonized dataset containing galaxies with complete physical measurements and human perceptual annotations was obtained. The merge based on the correspondence between OBJID and dr7objid yielded a final sample of 243,500 galaxies, which serves as the basis for all subsequent analyses. The dataset structure is summarized in Table 1.

Table 1. Final dataset variables after integration and cleaning (243,500 galaxies).

Category	Variable	Description
Identification	specobjid	Spectroscopic object identifier
Identification	OBJID	Photometric object identifier
Spatial coordinates	RA	Right Ascension (degrees)
Spatial coordinates	DEC	Declination (degrees)
Physical features	z	Redshift
Physical features	g_r	Color index (g − r)
Physical features	u_r	Color index (u − r)
Physical features	C_r	Concentration index (R90/R50)
Physical features	fracDeV_r	Fractional de Vaucouleurs component
Human perceptual features	t01_smooth_or_features_a01_smooth_debiased	Smooth morphology probability
Human perceptual features	t02_edgeon_a04_yes_debiased	Edge-on probability
Human perceptual features	t03_bar_a06_bar_debiased	Bar presence probability
Human perceptual features	t04_spiral_a08_spiral_debiased	Spiral structure probability
Human perceptual features	t05_bulge_prominence_a10_no_bulge_debiased	No bulge probability
Human perceptual features	t05_bulge_prominence_a11_just_noticeable_debiased	Just noticeable bulge probability
Human perceptual features	t05_bulge_prominence_a12_obvious_debiased	Obvious bulge probability
Human perceptual features	t05_bulge_prominence_a13_dominant_debiased	Dominant bulge probability
Human perceptual features	t07_rounded_a16_completely_round_debiased	Completely round probability
Human perceptual features	t07_rounded_a17_in_between_debiased	Intermediate shape probability
Human perceptual features	t07_rounded_a18_cigar_shaped_debiased	Cigar-shaped probability
Human perceptual features	t08_odd_feature_a19_ring_debiased	Ring probability

All selected variables are complete after preprocessing, with only a negligible number of missing spectroscopic identifiers. The resulting dataset defines two fully specified feature spaces, physical and perceptual, ensuring consistency for the fuzzy clustering, supervised validation, and interpretability analyses.

4.2. Exploratory Data Analysis

4.2.1. Physical Feature Space

Following the preprocessing and outlier filtering described in Section 3, an exploratory analysis of the physical variables was conducted prior to fuzzy modeling. The results are summarized in Figures 6–8.

Figure 6 presents the distributions of the redshift z and photometric color indices $g - r$ and $u - r$. The redshift spans approximately $0 < z < 0.25$, with a clear maximum around $z \approx 0.07$–0.10, corresponding to the region where the SDSS main galaxy sample is most complete. The decline at higher redshift reflects the expected behavior of a flux-limited survey.

The color index $g - r$ ranges approximately from -0.5 to 1.5, with a peak near $g - r \approx 0.8$, indicating the coexistence of blue, star-forming galaxies and red, passively evolving systems.

Similarly, $u - r$ spans roughly $0 < u - r < 4$, with a peak around $u - r \approx 2.7$, reflecting a predominance of evolved stellar populations while retaining a non-negligible fraction of actively star-forming galaxies.

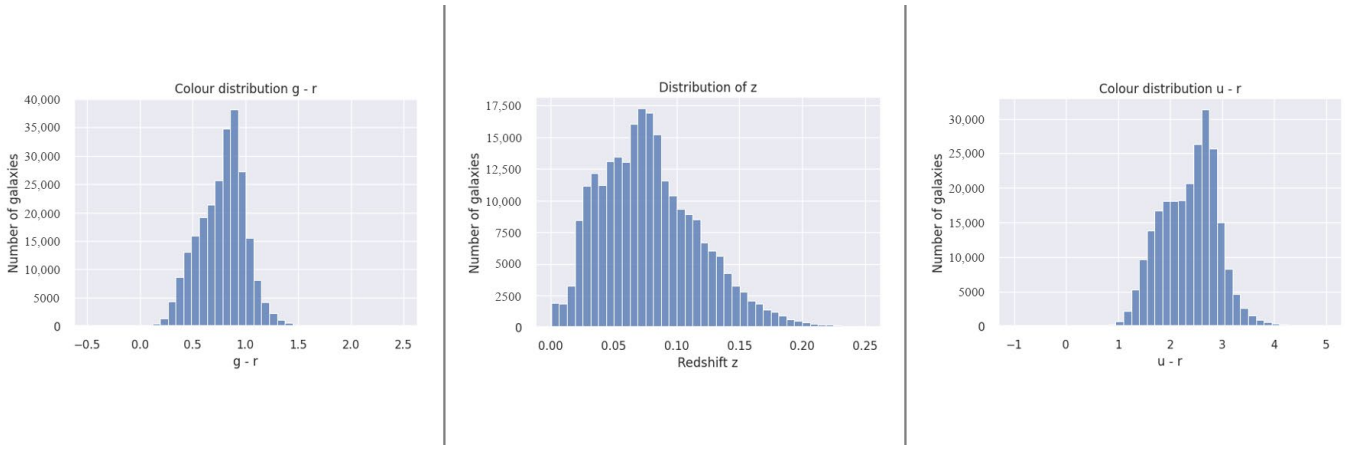

Figure 6. Distributions of redshift and photometric color indices (z, g − r, u − r).

Figure 7 displays the structural parameters C_r and $fracDeV_r$. The concentration index $C_r = R_{90}/R_{50}$ is primarily distributed within $2 < C_r < 3.5$, with a maximum around $C_r \approx 3$, distinguishing disk-dominated from centrally concentrated morphologies.

The variable $fracDeV_r$ exhibits a pronounced bimodal distribution, with peaks near 0 and 1. This indicates a clear separation between exponential disk profiles and bulge-dominated systems, with relatively few intermediate cases.

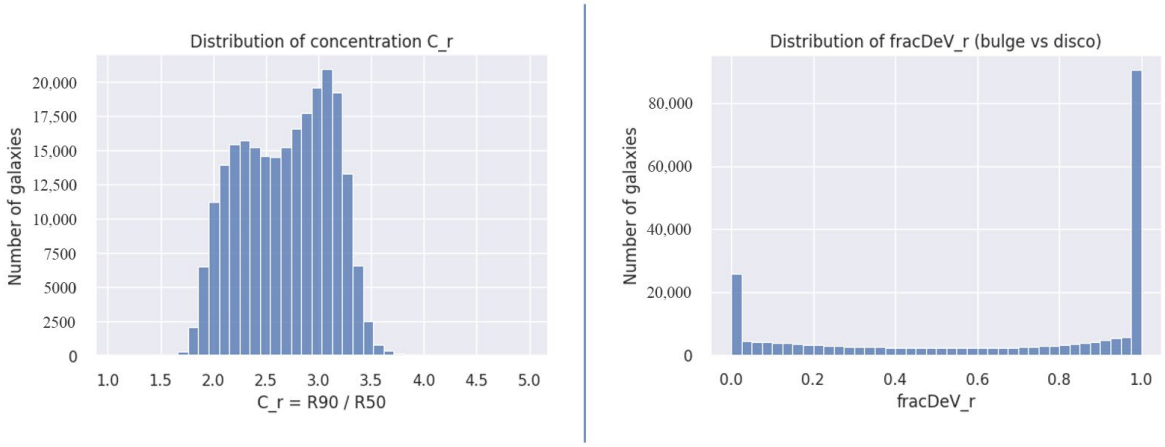

Figure 7. Distributions of structural parameters (C_r and fracDeV_r).

Figure 8 presents the Pearson correlation matrix among the physical variables. The color indices ($g - r$, $u - r$) are strongly correlated (≈0.87), reflecting their shared sensitivity to stellar populations. Both colors show moderate-to-strong correlations with structural parameters (C_r, fracDeV$_r$, ≈ 0.63–0.70), indicating a close association between redder populations and more concentrated morphologies.

The correlation between C_r and fracDeV$_r$ is particularly high (≈0.83), confirming structural consistency between independent morphological proxies. In contrast, redshift shows weak correlations with the remaining variables, as expected given its cosmological nature.

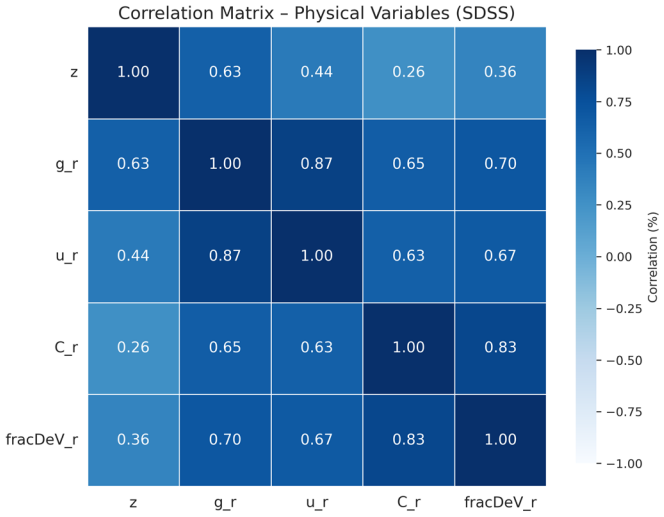

Figure 8. Pearson correlation matrix of physical variables.

Overall, the physical feature space exhibits coherent structural relationships with partial redundancy among morphology-related descriptors. This structure supports a separation between disk- and bulge-dominated regimes while preserving gradual transitions, making it well suited for fuzzy clustering.

4.2.2. Human Feature Space

An exploratory analysis was also conducted on the debiased Galaxy Zoo 2 probabilities prior to fuzzy clustering. The results are shown in Figures 9–11.

Figure 9 presents the distributions of key morphological probabilities. The bar probability is strongly concentrated near zero, indicating that clearly barred systems represent a minority of the sample. The smooth probability also peaks near zero, reflecting a predominance of non-smooth (typically disk-like) morphologies. The "no bulge" probability is similarly concentrated near zero, suggesting that most galaxies are perceived to contain some bulge component.

The spiral probability exhibits a polarized distribution, with a strong concentration near zero and a secondary accumulation toward unity, reflecting confidently identified spiral systems alongside non-spiral galaxies.

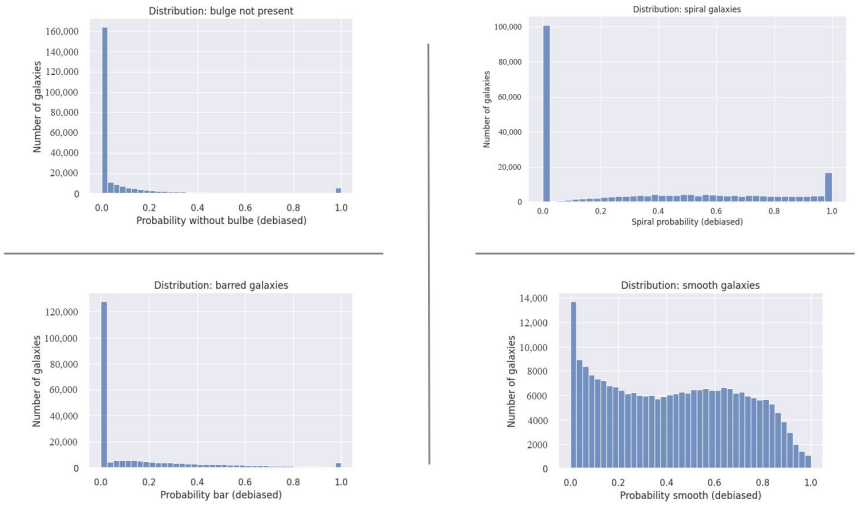

Figure 9. Distribution: barred, smooth, bulge not present, and spiral.

Figure 10 displays additional perceptual variables. Ring classifications are rare, with probabilities heavily skewed toward zero. The completely round variable shows a polarized distribution, reflecting a mixture of inclined disks and spheroidal systems. Cigar-shaped morphologies are uncommon, with a small high-probability tail. The dominant bulge probability is also strongly skewed toward zero, consistent with the limited fraction of clearly bulge-dominated systems.

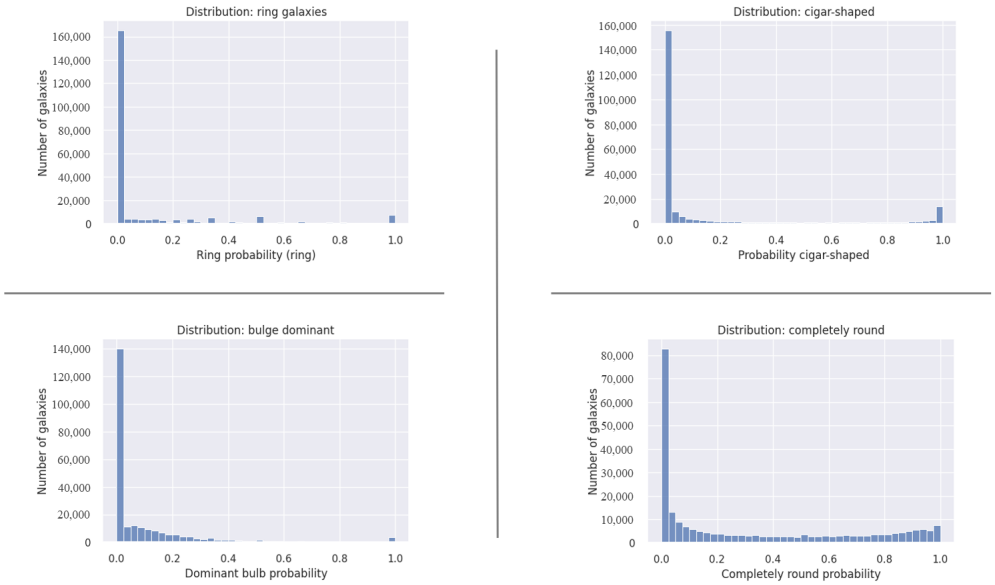

Figure 10. Distribution: ring, cigar-shaped, bulge dominant, and round.

Figure 11 presents the Pearson correlation matrix of the selected debiased probabilities. The matrix reveals consistent morphological relationships, including strong negative correlations between mutually exclusive categories (e.g., smooth vs. spiral) and positive correlations between compatible traits (e.g., smooth with round or bulge-dominant). Several weak correlations indicate complementary, rather than redundant, descriptors.

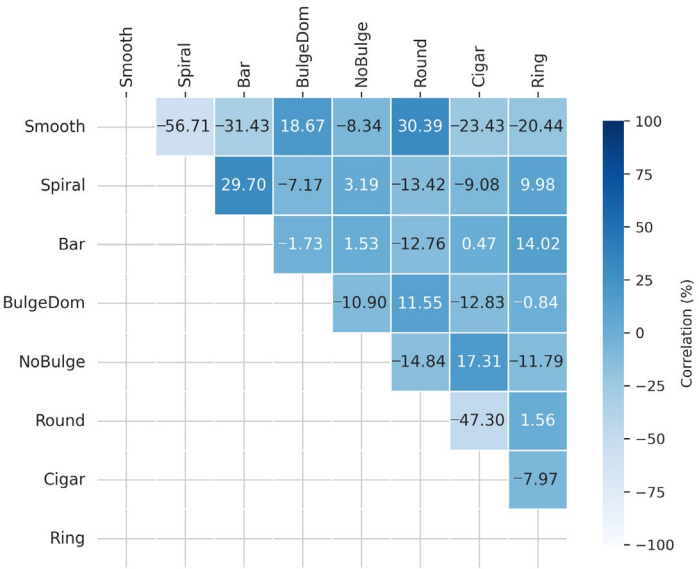

Figure 11. Pearson correlation matrix of selected debiased morphological probabilities.

These results confirm the internal consistency of Galaxy Zoo classifications and demonstrate that the probabilistic annotations provide a structured yet flexible representation of visual morphology. This makes them well suited for comparison with physically derived clustering regimes.

4.3. Fuzzy C-Means Segmentation

To identify latent morphological regimes, the Fuzzy C-Means (FCM) algorithm was applied independently to the normalized physical and perceptual feature spaces. Unlike hard clustering, FCM assigns graded memberships, allowing galaxies to occupy transitional regions between structural classes, an appropriate property given the continuous nature of galaxy morphology.

The number of clusters was determined using internal fuzzy validity indices, ensuring that segmentation was data-driven rather than imposed a priori.

4.3.1. Physical Feature Clusters

Cluster selection was guided by the Fuzzy Partition Coefficient (FPC) and the Xie–Beni index (XB). As shown in Table 2, the two-cluster configuration ($c = 2$) achieves the highest FPC (0.7597) and the lowest XB (0.1544), indicating optimal compactness and separation. Increasing the number of clusters produces a monotonic decrease in FPC and a marked increase in XB, particularly for $c \geq 5$, suggesting over-segmentation without structural gain. Both indices consistently support a two-cluster solution.

Table 2. Fuzzy cluster validity indices: Physical Features.

c	FPC	Xie-Beni
2	0.7597	0.1544
3	0.5940	0.5630
4	0.5109	0.5228
5	0.4357	1.2422
6	0.3862	1.6172

The centroids of the optimal configuration, expressed in the original variable scale, are reported in Table 3.

Table 3. Physical cluster centroids.

Cluster	z	g-r	u-r	C_r	fracDeV_r
0	0.0609	0.584	1.892	2.297	0.249
1	0.0914	0.923	2.718	2.977	0.919

A clear structural dichotomy emerges between the two clusters.

- **Cluster 0** exhibits bluer colors ($g - r = 0.58$, $u - r = 1.89$), lower concentration ($C_r = 2.30$), and a small de Vaucouleurs fraction (0.25), consistent with disk-dominated systems characterized by ongoing star formation.
- **Cluster 1** shows redder colors ($g - r = 0.92$, $u - r = 2.72$), higher concentration ($C_r = 2.98$), and a near-unity de Vaucouleurs fraction (0.92), indicating centrally concentrated spheroidal systems associated with evolved stellar populations.

Redshift differences between clusters are modest and likely reflect sampling effects rather than intrinsic morphological separation.

The coherence between photometric colors and structural parameters confirms that the fuzzy partition captures physically meaningful regimes. At the same time, membership degrees preserve intermediate cases that would be forced into rigid categories under hard clustering.

Figure 12 presents representative SDSS images from both clusters. Disk-dominated systems typically display elongated or structured morphologies with bluer colors, whereas bulge-dominated systems appear smoother, rounder, and redder. The figure also reports the corresponding membership degrees (μ_0, μ_1), where high values ($\mu \approx 0.9$) identify prototypical galaxies and intermediate values ($\mu \approx 0.4$–0.6) correspond to transitional systems with mixed disk–bulge characteristics.

Figure 12. Representative SDSS images of galaxies from the two physical fuzzy clusters.

Overall, the two-cluster solution recovers the well-known bimodal structure of galaxy populations. The consistency between validity indices (Table 2), centroid statistics (Table 3), and visual morphology (Figure 12) support the adequacy of the selected physical variables for capturing the dominant structural dichotomy in an unsupervised framework.

4.3.2. Human Feature Clusters

The FCM algorithm was subsequently applied to the human perceptual feature space derived from the debiased Galaxy Zoo 2 probabilities.

When all twelve perceptual variables were included, the clustering procedure did not reveal a stable partition. The FPC remained close to its theoretical lower bound ($\approx 1/c$), while the XB index exhibited extremely large values (on the order of 10^{10}–10^{11}) across all tested configurations ($c = 2 \ldots 6$). The results are summarized in Table 4.

Table 4. Fuzzy cluster validity indices (full human perceptual space).

c	FPC	Xie-Beni
2	0.5000	$\approx 0.5 \times 10^{10}$
3	0.3333	$\approx 1.9 \times 10^{10}$
4	0.2500	$\approx 4.5 \times 10^{10}$
5	0.2000	$\approx 8.1 \times 10^{11}$
6	0.1666	$\approx 5.8 \times 10^{11}$

Additionally, cluster centroids were nearly identical across configurations, differing only at high decimal precision. This behavior indicates weak compactness and separation in the high-dimensional perceptual space.

These results suggest that aggregated vote fractions form a highly smooth representation of morphology. Rather than discrete groupings, the perceptual space encodes gradual variation, consistent with the probabilistic aggregation underlying Galaxy Zoo classifications.

To isolate dominant perceptual dimensions, clustering was repeated using three primary variables: smooth probability, spiral probability, and edge-on probability. Under this reduced representation, clear structure emerges and validity indices improve substantially (Table 5).

Table 5. Fuzzy cluster validity indices for the reduced human perceptual feature space.

c	FPC	Xie–Beni
2	0.7288	0.2807
3	0.6928	0.1949
4	0.6271	0.4536
5	0.5788	0.7647
6	0.5500	0.8628

For $c = 2$, the indices indicate compact and well-separated partitions. The corresponding centroids are reported in Table 6.

Table 6. Centroids in the reduced human perceptual space ($c = 2$).

Cluster	Smooth	Spiral	Edge-On
Cluster A_0	0.2057	0.6532	0.1273
Cluster A_1	0.6248	0.0929	0.1240

The two-cluster solution separates spiral-dominated systems (A_0) from smooth, elliptical-like galaxies (A_1), corresponding to the classical late-type (disk-dominated) versus early type (bulge-dominated) distinction.

For $c = 3$, an additional cluster emerges characterized by high edge-on probability (Table 7).

Table 7. Centroids in the reduced human perceptual space (c = 3).

Cluster	Smooth	Spiral	Edge-On
Cluster B_0	0.3180	0.1208	0.7666
Cluster B_1	0.6590	0.0871	0.0448
Cluster B_2	0.1822	0.7176	0.0547

This three-cluster configuration differentiates smooth systems, spiral face-on systems, and spiral edge-on systems. Notably, the explicit separation of orientation effects in the perceptual space contrasts with the physical feature space, where inclination is not directly encoded.

For direct comparison with physical clustering, the two-cluster solution ($c = 2$) is retained as the primary perceptual partition.

4.3.3. Physical–Human Cluster Correspondence

To evaluate the relationship between structural (physical) and perceptual (human) segmentations, fuzzy memberships were converted into dominant assignments via maximum membership selection. The resulting contingency matrix is shown in Figure 13.

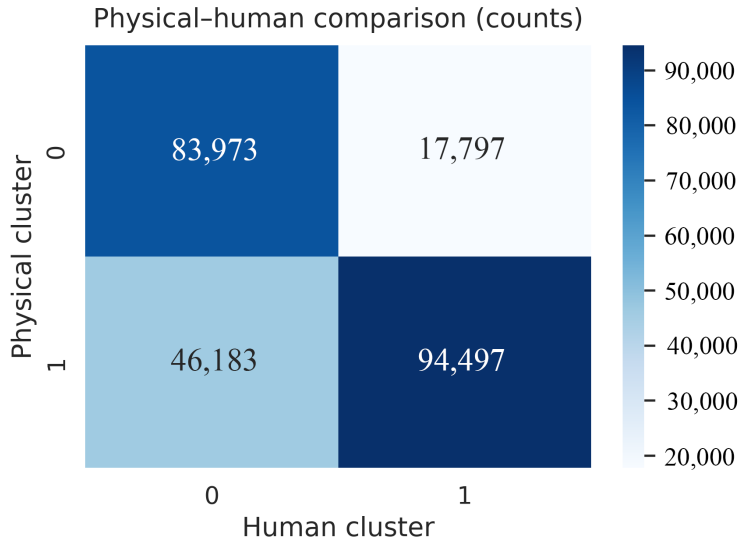

Figure 13. Contingency matrix between physical and human two-cluster solutions.

The comparison reveals substantial agreement between both representations. Specifically, 82.5% of galaxies assigned to physical cluster 0 are also classified into human cluster 0, while 67.2% of galaxies in physical cluster 1 correspond to human cluster 1.

Agreement is stronger for disk-dominated systems, suggesting that spiral structure and low concentration profiles are consistently identified by both physical parameters and visual inspection. The lower agreement for bulge-dominated systems likely reflects orientation effects and the presence of intermediate morphologies.

As observed in Section 4.3.2, the perceptual space explicitly isolates edge-on galaxies when $c = 3$, whereas the physical feature space does not encode inclination. This discrepancy provides a plausible explanation for residual disagreement between partitions.

Table 8 presents representative galaxies with dominant cluster assignments and fuzzy membership degrees across both physical and perceptual spaces.

Table 8. Representative galaxies with cluster labels and fuzzy memberships in physical (c = 2) and human spaces (c = 2, 3).

specobjid	c_phys	c_phys_0	c_phys_1	c_h2	c_h2_0	c_h2_1	c_h3	c_h3_0	c_h3_1	c_h3_2
327795784667916288	0	0.909326	0.090674	0	0.541977	0.458023	2	0.192059	0.380568	0.427373
384079506022885376	0	0.593500	0.406500	1	0.345340	0.654660	0	0.840389	0.102390	0.057221
1825148676935280640	1	0.003439	0.996561	1	0.046374	0.953626	1	0.009494	0.978013	0.012493
2046957849925412864	1	0.244207	0.755793	0	0.611142	0.388858	2	0.083417	0.376473	0.540110
2034627920424626176	0	0.891380	0.108620	0	0.518786	0.481214	0	0.813961	0.089651	0.096388
2651588077599877120	1	0.429393	0.570607	1	0.081087	0.918913	1	0.033256	0.926815	0.039929
2991654116721190912	0	0.923146	0.076854	0	0.785650	0.214350	2	0.062267	0.214990	0.722742
872695886486988800	0	0.856760	0.143240	1	0.252228	0.747772	1	0.066826	0.740180	0.192995
1523421240763639808	1	0.043213	0.956787	1	0.073138	0.926862	1	0.035150	0.912305	0.052545
1610149342214121472	1	0.019801	0.980199	1	0.043611	0.956389	1	0.009037	0.978720	0.012243

Rather than binary assignments, galaxies exhibit graded memberships, reflecting transitional morphologies and structural ambiguity. The fuzzy framework therefore captures gradual variation that would be obscured under hard classification.

Overall, the alignment between physical and perceptual partitions indicates that the selected structural parameters recover, to a substantial extent, the primary morphological dichotomy identified by human classifiers. At the same time, systematic discrepancies highlight the role of perceptual and orientation-dependent effects not fully captured by physical descriptors.

4.4. Supervised Cross-Space Prediction Using XGBoost

Following the fuzzy segmentation of the physical and perceptual feature spaces, supervised models were trained to quantify cross-space predictability. Extreme Gradient Boosting (XGBoost) classifiers were fitted using the physical variables $z, g-r, u-r, C_r, fracDeV_r$ as predictors.

Two prediction tasks were defined:

- **Physical self-consistency test**: prediction of physical cluster labels from physical variables.
- **Cross-space prediction**: prediction of human cluster labels ($c = 2$) from the same physical variables.

Fuzzy memberships were converted into dominant labels via maximum membership selection. Stratified 80/20 train–test splits were applied to preserve class balance, providing a standard and balanced setting for training and evaluation.

4.4.1. Prediction of Physical Clusters from Physical Variables

As an internal benchmark, XGBoost was trained to reproduce the physical cluster assignments derived from the FCM segmentation.

The classifier achieved an accuracy of 99.54% on the test set (Figure 14), with balanced precision and recall across both classes. The confusion matrix shows 225 misclassifications out of 48,490 test galaxies.

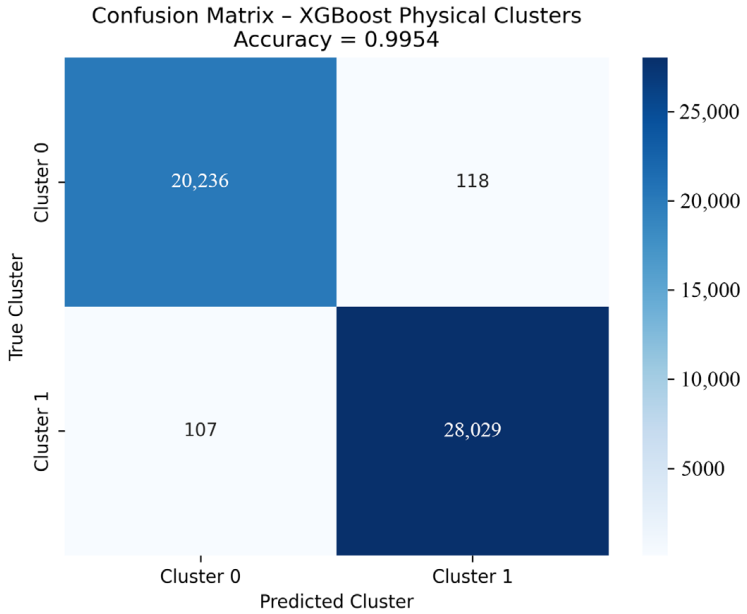

Figure 14. Confusion Matrix–XGBoost Physical Clusters.

These results indicate that the two physical clusters are highly separable within the original feature space. The near-perfect predictive performance confirms that the fuzzy partition corresponds to a well-defined structural boundary rather than a diffuse or unstable segmentation.

4.4.2. Prediction of Human Clusters from Physical Variables

The second experiment evaluated whether structural photometric parameters could approximate the perceptual clustering derived from Galaxy Zoo classifications.

When predicting the human two-cluster solution ($c = 2$) using only physical variables, XGBoost achieved an accuracy of 75.36% (Figure 15), with a macro-average F1-score of approximately 0.75.

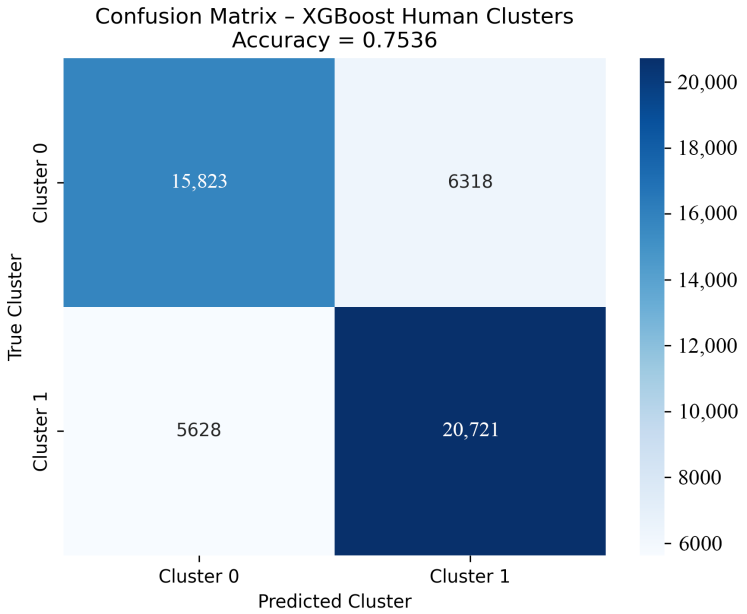

Figure 15. Confusion Matrix–XGBoost Human Clusters.

Class-wise performance remained balanced:

- Precision: 0.738 (cluster 0), 0.766 (cluster 1).
- Recall: 0.715 (cluster 0), 0.786 (cluster 1).

The observed performance indicates that a substantial portion of perceptual morphological structure is recoverable from measurable physical parameters. However, the remaining discrepancy suggests that human classification incorporates additional factors not fully captured by the selected descriptors. These likely include orientation effects, fine-scale visual features, and nonlinear perceptual cues.

4.5. Explainable Analysis (Global and Local SHAP Interpretation)

To interpret the decision structure of the supervised models, SHAP was applied to both the physical-cluster classifier and the human-cluster classifier. Since both models use the same five physical variables ($z, g-r, u-r, C_r, fracDeV_r$), differences in attribution reflect differences in learned decision boundaries rather than differences in feature availability.

4.5.1. Global Interpretation

The SHAP summary and average importance plots for the physical model (Figure 16) reveal a clear hierarchy of feature influence. The dominant driver is $fracDeV_r$, followed by C_r, while the color indices ($u-r$, $g-r$) provide secondary but consistent contributions. Redshift has comparatively minor impact.

High values of $fracDeV_r$ and C_r strongly increase the probability of belonging to the bulge-dominated cluster, indicating that the physical segmentation is primarily governed by structural concentration and bulge prominence.

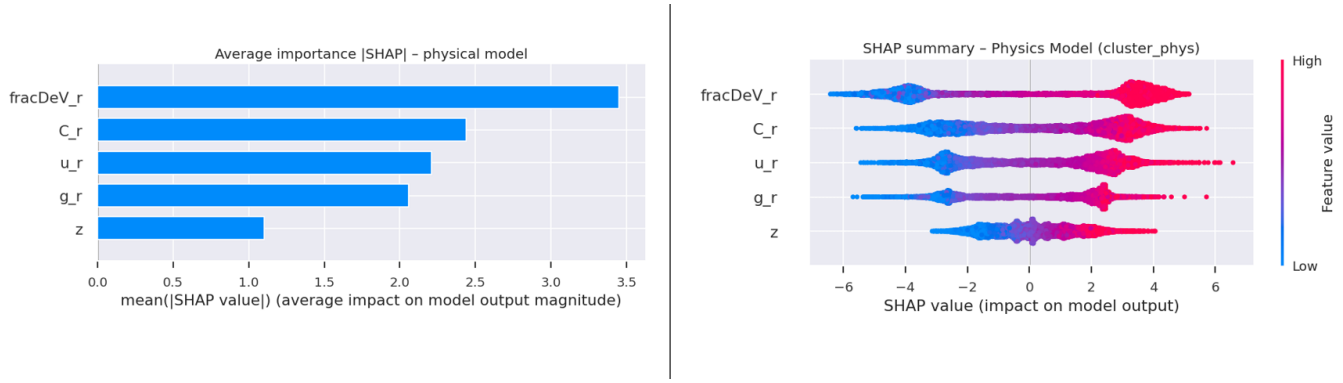

Figure 16. SHAP Summary and Average Importance: Physical Model.

In contrast, the human-cluster model (Figure 17) exhibits a different ordering. The dominant variable becomes $g - r$, followed by C_r, while $fracDeV_r$ has reduced relative influence. This shift indicates that perceptual clustering places greater weight on chromatic contrast and visually salient morphology than on detailed structural decomposition parameters. The modest but non-negligible contribution of redshift likely reflects observational or sampling effects rather than intrinsic morphological information.

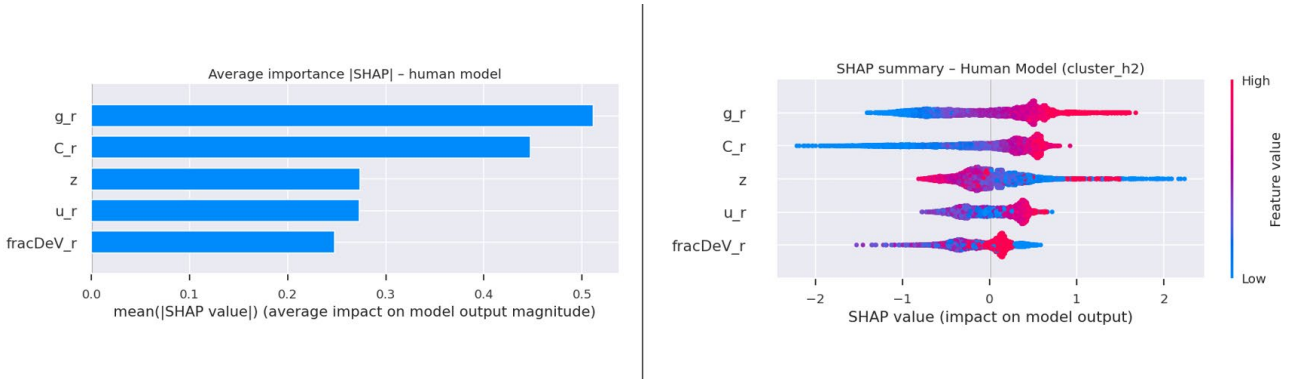

Figure 17. SHAP Summary and Average Importance: Human Model.

From an astronomical perspective, this difference is physically coherent. In the physical model, fracDeV$_r$ and C_r dominate because both variables are directly linked to the integrated structural organization of the galaxy, particularly bulge–disk dominance and central light concentration. In the human-cluster model, however, color becomes more important because human observers respond first to the global visual appearance of the system. Redder galaxies are more often perceived as smooth, evolved, and bulge-dominated, whereas bluer galaxies are more readily associated with spiral or feature-rich morphologies. The concentration index C_r remains relevant because the distribution of light also affects visual compactness, but parameters such as fracDeV$_r$ are not directly perceived in the same explicit way; rather, human classifiers capture the visible consequence of those structural properties.

The global SHAP comparison shows that both models capture the same broad morphological separation, red, concentrated systems versus blue, extended systems, while assigning different relative importance to structural and photometric features.

4.5.2. Local Interpretation

Local SHAP explanations (Figures 18 and 19) provide instance-level insight into how physical variables influence predicted cluster membership. Both models receive identical inputs; the difference lies in the target labels (physical versus human-derived clusters).

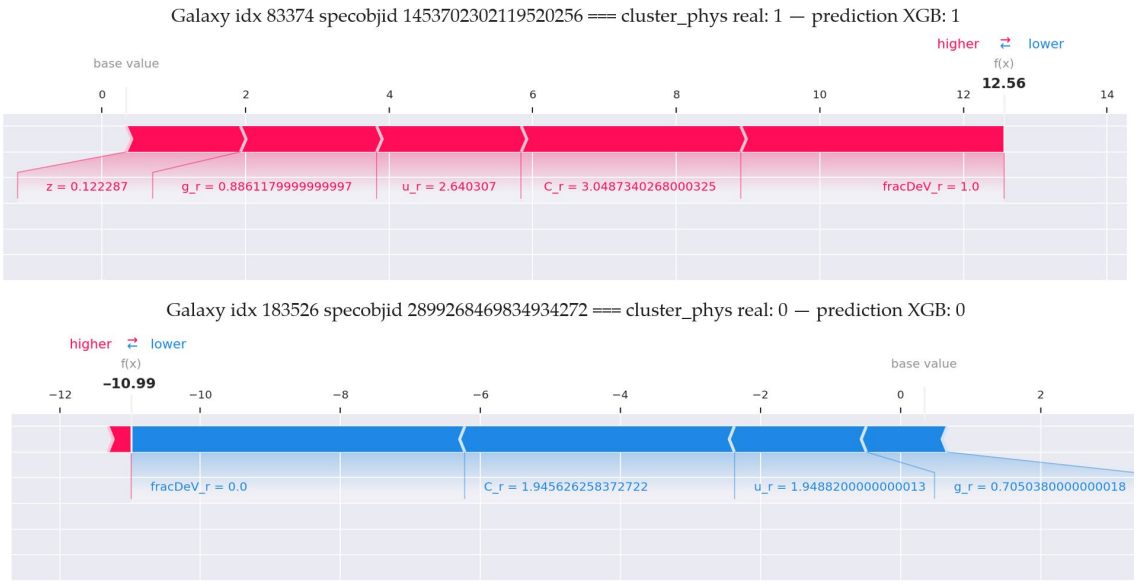

Figure 18. Local SHAP explanation for the physical cluster prediction (XGBoost model).

Figure 19. Local SHAP explanation for the human-derived cluster prediction (XGBoost model).

For galaxy specobjid 1453702302119520256, predicted as cluster 1 in both models, the dominant positive contributions arise from high $fracDeV_r = 1.0$, elevated concentration ($C_r \approx 3.05$), and red colors ($g - r \approx 0.89$, $u - r \approx 2.64$). These features jointly drive the prediction toward the bulge-dominated regime.

Conversely, for galaxy specobjid 2899268469834934272, predicted as cluster 0 in both models, low concentration ($C_r \approx 1.95$), $fracDeV_r = 0.0$, and bluer colors ($g - r \approx 0.70$, $u - r \approx 1.95$) shift the prediction toward the disk-dominated class.

The SHAP decompositions are qualitatively consistent across both prediction tasks, indicating that overlapping physical signals underlie both structural and perceptual classifications. However, differences in the magnitude of contributions reflect the distinct weighting of features observed in the global analysis.

These instance-level explanations reinforce the broader finding that measurable structural parameters encode a substantial portion of the morphological information captured

by human classification, while also revealing systematic differences in how these signals are prioritized.

4.6. Visual and Interactive Validation Results

While quantitative metrics, clustering validity indices, and explainability analyses provide strong statistical support for the proposed framework, visual inspection remains essential in astrophysical morphology studies. To further validate the coherence between structural (physical) and perceptual (human) segmentations, a two-stage validation strategy is introduced.

First, a structured visual comparison of representative galaxies grouped by agreement and disagreement between physical and human cluster assignments enables direct morphological interpretation of concordant and discordant cases. Second, an interactive exploration environment supports dynamic inspection at the individual-galaxy level, linking physical descriptors, perceptual annotations, and clustering outcomes.

4.6.1. Visual Cluster Correspondence Analysis

To examine correspondence patterns, galaxies were extracted from the physical–human contingency space and displayed as SDSS cutouts. Two categories were considered: agreement cases, including the (1,1) and (0,0) groups (Figures 20 and 21), and disagreement cases, including the (1,0) and (0,1) groups (Figures 22 and 23).

In the (1,1) group (phys = 1, hum = 1), galaxies typically exhibit centrally concentrated, smooth light distributions with reddish tones and minimal visible spiral structure. These properties are consistent with bulge-dominated systems and early type perceptual classifications.

In the (0,0) group (phys = 0, hum = 0), galaxies are predominantly extended disk systems, including both face-on and edge-on configurations, often displaying visible structure and comparatively blue colors. These objects align with late-type regimes in both physical and perceptual representations.

The disagreement groups reveal the origin of residual discrepancies. In the (1,0) group (phys = 1, hum = 0), many galaxies exhibit disk-like morphologies, frequently edge-on or dust-affected systems, whose integrated photometric properties (redder colors and higher concentration) resemble early type objects. Projection and attenuation effects therefore bias structural indicators relative to visual interpretation.

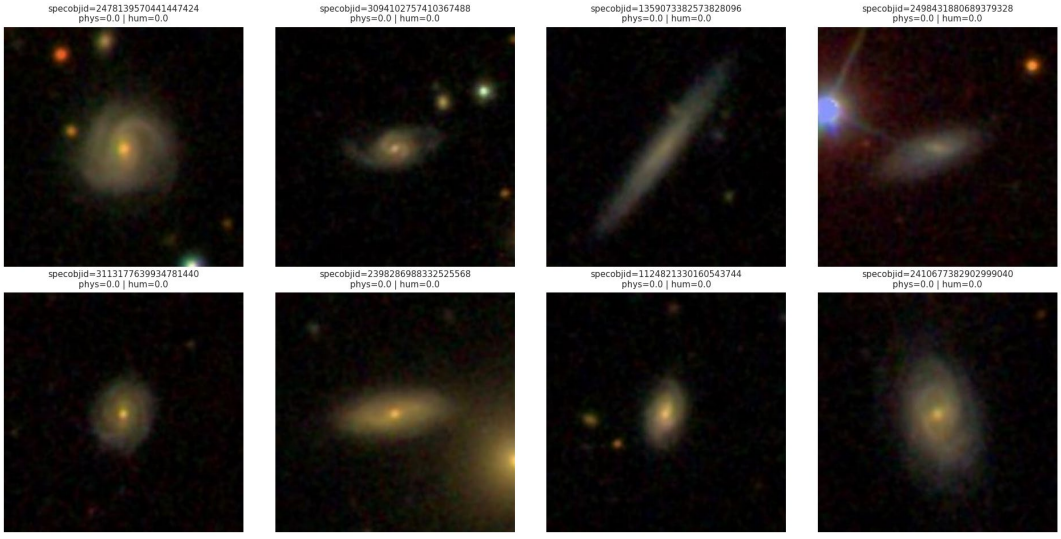

Figure 20. SDSS cutouts for agreement cases between physical and human clusters (0–0).

Figure 21. SDSS cutouts for agreement cases between physical and human clusters (1–1).

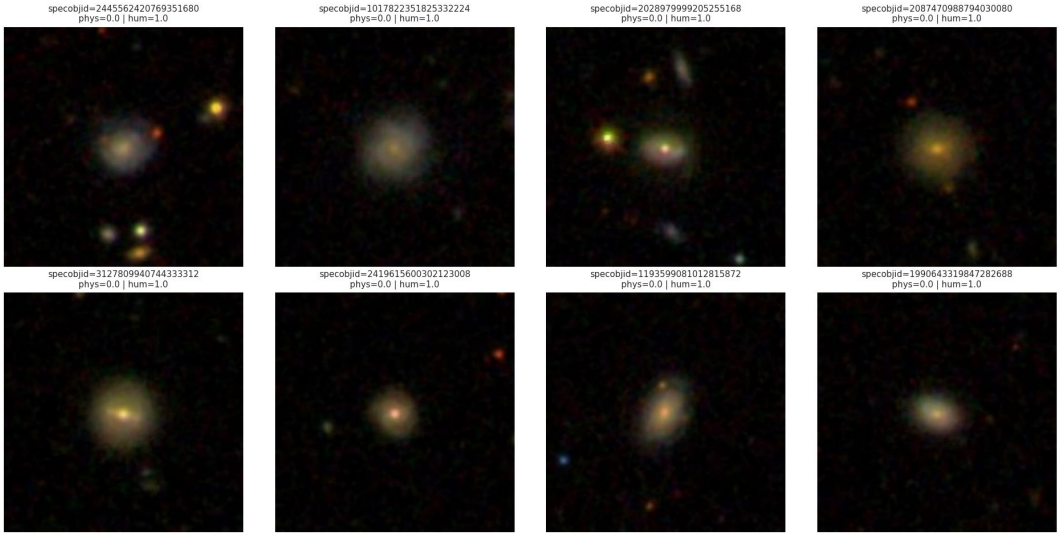

Figure 22. SDSS cutouts for disagreement cases between physical and human clusters (0–1).

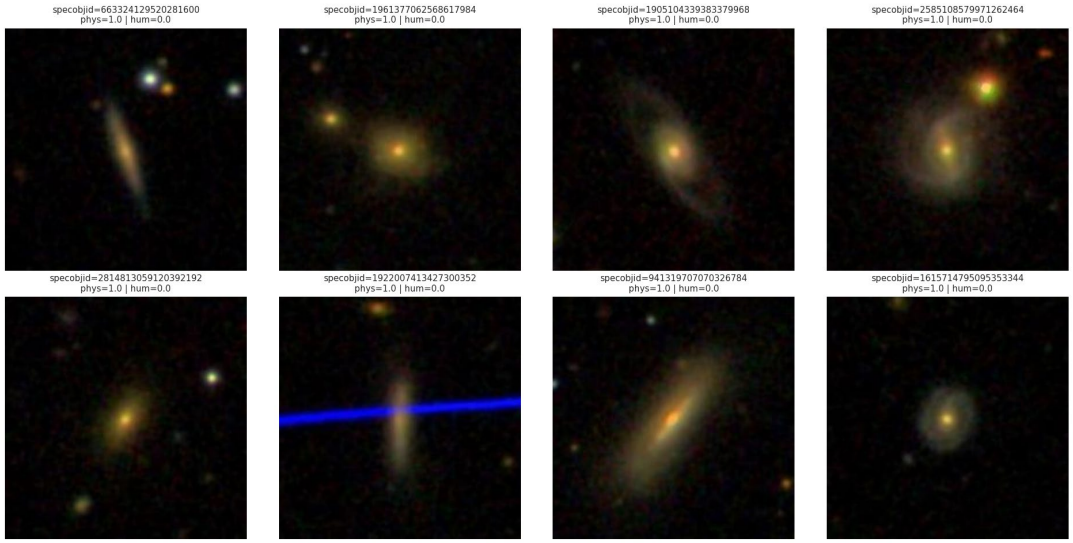

Figure 23. SDSS cutouts for disagreement cases between physical and human clusters (1–0).

Conversely, in the (0,1) group (phys = 0, hum = 1), galaxies often appear visually smooth but retain structural or photometric properties consistent with disk-dominated systems. These cases likely reflect limited angular resolution, low-contrast spiral features, or transitional morphologies.

There is greater consistency in early and late-type prototypical systems, while discrepancies are concentrated in transitional and orientation-sensitive regimes. These visual patterns are consistent with the SHAP analysis, which demonstrated that both representations are based on similar physical cues, but differ in their relative weighting and in their sensitivity to appearance-related cues.

4.6.2. Interactive Morphological and Structural Exploration

To complement static visual comparison, an interactive exploration environment was implemented to enable dynamic inspection of the galaxy sample. Objects are displayed in celestial coordinates (RA, DEC) and color-coded according to physical cluster assignment. Selecting a galaxy retrieves its SDSS image and displays its structural parameters together with the corresponding Galaxy Zoo vote fractions.

This interface enables simultaneous inspection of:

- Physical descriptors $(z, g-r, u-r, C\_r, fracDeV\_r)$;
- Unsupervised physical cluster assignment;
- Human perceptual classification;
- Detailed vote distributions.

The examples shown in Figures 24 and 25 illustrate both concordant (1,1), (0,0) and discordant (1,0), (0,1) regimes. In agreement cases, structural parameters, clustering results, and visual morphology form a consistent interpretation. In disagreement cases, the interface allows immediate identification of orientation effects, dust reddening, compact transitional morphologies, or ambiguous visual features.

Rather than constituting a separate analytical step, the interactive environment functions as a traceability layer that links data-driven segmentation with direct morphological inspection. It provides transparency at the object level and supports qualitative validation of statistical findings, reinforcing the interpretability of the proposed framework. The processed dataset and the interactive notebook supporting this object-level exploration are provided in the Supplementary Materials.

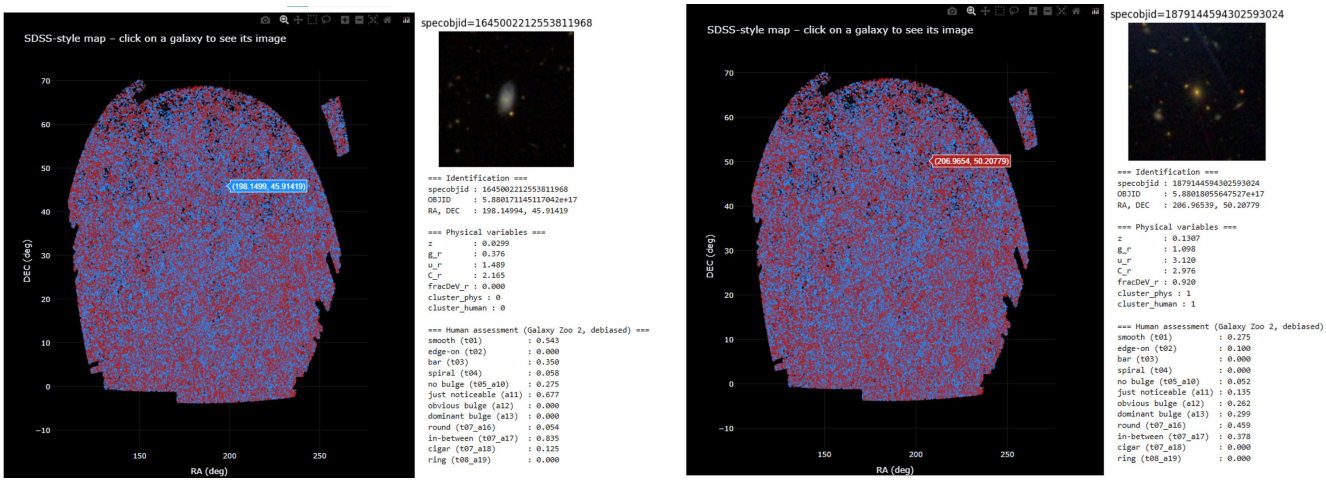

Figure 24. Visual examples of concordant physical and human classifications (1–1 and 0–0 regimes).

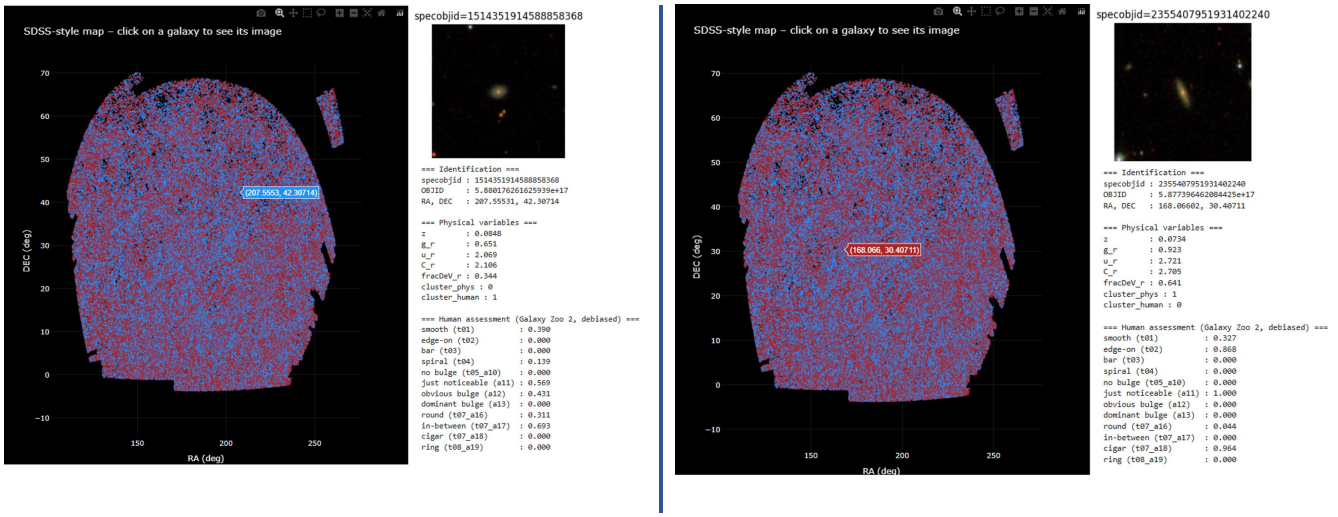

Figure 25. Visual examples of discordant physical and human classifications (1–0 and 0–1 regimes).

5. Discussion and Future Work

5.1. Discussion

The results obtained in this study indicate a substantial degree of coherence between structural photometric parameters and human perceptual classifications. However, several methodological and conceptual aspects warrant further consideration.

First, the physical feature space was intentionally restricted to five primary variables ($z, g-r, u-r, C_r, fracDeV_r$). While these parameters capture key aspects of stellar populations and structural concentration, they do not exhaust the morphological information available in contemporary surveys. Additional descriptors, such as non-parametric morphological indicators (e.g., asymmetry, Gini coefficient, M20), spectroscopic diagnostics (e.g., star formation rate, metallicity), or kinematic measurements were not included. Consequently, physical clustering reflects a reduced structural subspace rather than a fully comprehensive morphological representation.

Second, the selection of $c = 2$ emphasizes the classical early–late dichotomy. Although this choice is supported by validity indices and interpretability, galaxy morphology is inherently multi-dimensional. The exploratory analysis with $c = 3$ in the perceptual space revealed orientation-driven segregation (edge-on systems), suggesting that additional physically meaningful regimes may exist beyond the binary partition. A systematic evaluation of cluster stability across different values of c could further reveal substructures such as S0 galaxies, barred spirals, or compact transitional systems.

Third, the supervised stage relies on hard labels derived from fuzzy memberships. While this enables classification and SHAP-based interpretation, it necessarily discards part of the uncertainty encoded in the membership matrix. Future extensions could incorporate fuzzy memberships directly into the supervised objective, preserving transitional states more faithfully.

Fourth, discrepancies between physical and human clusters highlight the influence of projection effects, dust attenuation, and image resolution. The physical feature space does not explicitly encode inclination, whereas human observers are highly sensitive to projected morphology. As a result, some disagreements reflect differences in representational dimensionality rather than classification errors.

Finally, although Galaxy Zoo probabilities were corrected for known biases, perceptual classifications remain influenced by survey depth, observational conditions, and collective voting dynamics. The perceptual space should therefore be interpreted as an aggregated cognitive signal rather than a definitive morphological ground truth.

Taken together, these observations indicate that the proposed framework should not be interpreted as defining a definitive morphological taxonomy. Instead, it provides a structured methodology for comparing complementary representations of galaxy morphology and for systematically analyzing their agreement, divergence, and transitional regimes.

Beyond its specific application to galaxy morphology, the broader contribution of this study lies in the proposal of an interpretable framework for relating physical measurements, human perception, and AI-based analysis within a unified analytical setting. In astronomy, as in many other domains, working with data involves prediction, interpretation, and learning. The framework developed here makes it possible to examine how measurable structure, collective human judgment, and explainable machine learning interact, supporting a more transparent reading of the data. In this sense, its value lies in identifying agreement between physical and perceptual classifications and in turning disagreement into interpretable and analytically meaningful information.

5.2. Future Work

From an astronomical perspective, future research may incorporate a broader range of physical descriptors, including non-parametric morphological indices (e.g., asymmetry, Gini coefficient, M20), spectroscopic diagnostics, stellar mass estimates, and environmental indicators. Expanding the feature space would enable evaluation of the robustness of the physical–perceptual correspondence when additional structural and evolutionary information is considered. Similarly, exploring alternative values of c and systematically assessing cluster stability may reveal substructures beyond the classical early–late dichotomy.

Methodologically, an important extension involves integrating fuzzy memberships directly into supervised learning, avoiding hard assignment simplifications and enabling a more faithful representation of transitional galaxies. In addition, comparing tree-based explainable models with interpretable deep learning approaches may clarify how different modeling paradigms capture structural signals and decision boundaries.

A particularly promising direction lies in representation learning methods capable of capturing morphological patterns that are not fully reflected in aggregated tabular variables. Vision transformer architectures applied directly to galaxy images may enable detection of visual features such as spiral structure, bars, inclination effects, and more subtle morphological characteristics. In this context, multimodal approaches combining structural parameters with imaging data could provide a more complete and unified representation of galaxy morphology, bridging the gap between physical descriptors and human perceptual interpretation.

Beyond astronomy, the proposed framework is extensible to domains in which automated systems and human judgment coexist. In medical imaging, for example, machine learning models extract statistical patterns from pixel-level data, while clinicians rely on perceptual and experiential cues. A structured comparison between data-driven segmentation and expert perception, supported by explainable AI, may help identify systematic agreement and disagreement regimes in diagnostic processes. Similar principles may apply in radiology, pathology, and other image-based decision contexts.

More broadly, future research may formalize the interaction between automated pattern recognition and human interpretation, not only to improve predictive performance but also to better understand the conditions under which algorithmic and perceptual reasoning converge or diverge. As a more exploratory direction, alternative clustering paradigms, including quantum-inspired approaches such as quantum k-means, could be evaluated to assess whether non-classical representations enable the identification of novel grouping structures. Given the current limitations in interpretability for such models, their

analysis may rely on surrogate modeling techniques combined with post hoc explanation methods such as SHAP or LIME.

6. Conclusions

This study introduced a structured framework (FAS-XAI) for comparing physically derived and human-based classifications of galaxy morphology through the integration of fuzzy clustering, supervised modeling, and explainable artificial intelligence.

Using a reduced but physically meaningful set of photometric parameters, the physical clustering recovered the dominant early–late morphological dichotomy observed by human classifiers. The strong agreement in prototypical systems indicates that global color and structural concentration encode a substantial fraction of the information used in visual classification. At the same time, systematic divergences in transitional, inclined, or low-contrast systems indicate that structural photometry and human perception capture complementary aspects of galaxy morphology.

The incorporation of XGBoost and SHAP enabled identification of the structural drivers underlying both physical and human-derived clusters. Although both representations rely on the same physical variables, differences in their relative importance reveal distinct sensitivities to structural concentration and visually salient features. Importantly, the proposed framework makes explicit not only where classifications coincide, but also where they diverge and why.

While the present analysis focuses on a limited feature set and a two-cluster configuration, the methodology is modular and extensible. Its main contribution lies not in proposing a definitive morphological taxonomy, but in providing a transparent and reproducible approach for analyzing agreement, discrepancy, and transitional regimes between physical measurements, human perception, and interpretable machine learning. Beyond the specific astronomical application, the proposed methodology has strong potential for domains in which physical measurements, human interpretation, and AI-assisted decision-making need to be integrated in an interpretable and transparent way.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/ai7050159/s1. The repository includes the processed dataset (*df_clean.csv*) and an interactive Python notebook (*Interactive Galaxy Images.ipynb*) that enables object-level exploration of the galaxy sample, linking SDSS imagery with physical features, clustering assignments, and perceptual annotations.

Author Contributions: Conceptualization, G.M.D.; methodology, G.M.D.; software, G.M.D.; validation, G.M.D., A.M.R.-R. and E.M.A.N.; formal analysis, G.M.D.; investigation, G.M.D.; resources, G.M.D.; data curation, G.M.D.; writing—original draft preparation, G.M.D.; writing—review and editing, G.M.D.; visualization, G.M.D.; supervision, G.M.D., A.M.R.-R. and E.M.A.N.; project administration, G.M.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable. The study did not involve human participants directly; it used anonymized data and public annotations from the Galaxy Zoo dataset.

Data Availability Statement: Publicly available datasets were used in this study. Galaxy Zoo 2 data are accessible at https://data.galaxyzoo.org/, and Sloan Digital Sky Survey Data Release 17 (SDSS DR17) data are available at https://www.sdss4.org/dr17/ (accessed on 5 February 2026). The processed dataset (*df_clean.csv*), containing integrated physical and perceptual variables used in this study, together with an interactive notebook for visual exploration and qualitative validation, are available at: http://doi.org/10.3390/ai7050159.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zooniverse. Galaxy Zoo. 2004. Available online: https://www.zooniverse.org/projects/zookeeper/galaxy-zoo (accessed on 15 February 2026).
2. Lintott, C.J.; Schawinski, K.; Slosar, A.; Land, K.; Bamford, S.; Thomas, D.; Raddick, M.J.; Nichol, R.C.; Szalay, A.; Andreescu, D.; et al. Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Mon. Not. R. Astron. Soc.* **2008**, *389*, 1179–1189. [CrossRef]
3. Molnar, C. Interpretable Machine Learning. In *A Guide for Making Black Box Models Explainable*; Lulu: Durham, NC, USA, 2019. Available online: https://christophm.github.io/interpretable-ml-book (accessed on 10 March 2026).
4. Medina-Rosales, E.; Cabrera-Vives, G.; Miller, C.J. Mitigating Bias in Deep Learning: Training Unbiased Models on Biased Data for the Morphological Classification of Galaxies. *arXiv* **2023**, arXiv:2308.11007. [CrossRef]
5. Hackstein, S.; Kinakh, V.; Bailer, C.; Melchior, M. Evaluation metrics for galaxy image generators. *Astron. Comput.* **2023**, *42*, 100685. [CrossRef]
6. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The Fuzzy C-Means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [CrossRef]
7. Doshi-Velez, F.; Kim, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv* **2017**, arXiv:1702.08608. [CrossRef]
8. Ginsburg, A.; Sipőcz, B.M.; Brasseur, C.E.; Cowperthwaite, P.S.; Craig, M.W.; Deil, C.; Groener, A.M.; Guillochon, J.; Guzman, G.; Liedtke, S.; et al. Astroquery: An Astronomical Web-querying Package in Python. *Astron. J.* **2019**, *157*, 98. [CrossRef]
9. D'Isanto, A.; Polsterer, K.L. Photometric redshift estimation via deep learning Generalized and pre-classification-less, image based, fully probabilistic redshifts. *Astron. Astrophys.* **2018**, *609*, A111. [CrossRef]
10. Hoyle, B. Measuring photometric redshifts using galaxy images and Deep Neural Networks. *Astron. Comput.* **2016**, *16*, 34–40. [CrossRef]
11. Caldeira, J.; Wu, W.K.; Nord, B.; Avestruz, C.; Trivedi, S.; Story, K.T. DeepCMB: Lensing reconstruction of the cosmic microwave background with deep neural networks. *Astron. Comput.* **2019**, *28*, 100307. [CrossRef]
12. Barra, V.; Delouille, V.; Kretzschmar, M.; Hochedez, J.F. Fast and robust segmentation of solar EUV images: Algorithm and results for solar cycle 23. *Astron. Astrophys.* **2009**, *505*, 361–371. [CrossRef]
13. Marín Díaz, G.; Gómez Medina, R.; Aijón Jiménez, J.A. Integrating Fuzzy C-Means Clustering and Explainable AI for Robust Galaxy Classification. *Mathematics* **2024**, *12*, 2797. [CrossRef]
14. Pickens, A.; Sengupta, S. Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means, Fuzzy C-Means and Evolutionary Neural Networks. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 695–719. [CrossRef]
15. Takeuchi, T.T.; Cooray, S.; Kano, R.R. Galaxy Evolution with Manifold Learning. *Entropy* **2026**, *28*, 288. [CrossRef] [PubMed]
16. Reza, M. Galaxy morphology classification using automated machine learning. *Astron. Comput.* **2021**, *37*, 100492. [CrossRef]
17. Yao, T.; Sun, L.; Geng, L.; Xu, Y.; Xu, Z.; Hu, K.; Chen, X.; Liao, P.; Wang, J. Exploring Optimisation Pathways for Underground Space Quality Under the Synergy of Multidimensional Perception and Environmental Parameters. *Buildings* **2025**, *15*, 204. [CrossRef]
18. Zhang, Y.; Liang, B.; Chen, B.; Torrens, P.M.; Atashzar, S.F.; Lin, D.; Sun, Q. Force-Aware Interface via Electromyography for Natural VR/AR Interaction. *ACM Trans. Graph.* **2022**, *41*, 268. [CrossRef]
19. Vallejo, M.; De La Espriella, C.; Gómez-Santamaría, J.; Ramírez-Barrera, A.F.; Delgado-Trejos, E. Soft metrology based on machine learning: A. review. *Meas. Sci. Technol.* **2019**, *31*, 32001. [CrossRef]
20. Marín Díaz, G. A Unified Fuzzy–Explainable AI Framework (FAS-XAI) for Customer Service Value Prediction and Strategic Decision-Making. *Ai* **2025**, *7*, 3. [CrossRef]
21. Willett, K.W.; Lintott, C.J.; Bamford, S.P.; Masters, K.L.; Simmons, B.D.; Casteels, K.R.; Edmondson, E.M.; Fortson, L.F.; Kaviraj, S.; Keel, W.C.; et al. Galaxy zoo 2: Detailed morphological classifications for 304 122 galaxies from the sloan digital sky survey. *Mon. Not. R. Astron. Soc.* **2013**, *435*, 2835–2860. [CrossRef]
22. Verma, R.K.; Tiwari, R.; Thakur, P.S. Thakur, Partition Coefficient and Partition Entropy in Fuzzy C Means Clustering. *J. Sci. Res. Rep.* **2023**, *29*, 1–6. [CrossRef]
23. Ghosh, A.; Mishra, N.S.; Ghosh, S. Fuzzy clustering algorithms for unsupervised change detection in remote sensing images. *Inf. Sci.* **2011**, *181*, 699–715. [CrossRef]
24. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016), San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
25. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4766–4775.
26. Sirapangi, M.; Gopikrishnan, S. MAIPFE: An Efficient Multimodal Approach Integrating Pre-Emptive Analysis, Personalized Feature Selection, and Explainable AI. *Comput. Mater. Contin.* **2024**, *79*, 2229–2251. [CrossRef]
27. Lundberg, S.M.; Erion, G.G.; Lee, S.I. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**, arXiv:1802.03888.

28. Ullah, I.; Rios, A.; Gala, V.; Mckeever, S. Explaining deep learning models for tabular data using layer-wise relevance propagation. *Appl. Sci.* **2022**, *12*, 136. [CrossRef]
29. Chen, W.H.; Hsieh, B.C.; Hsu, C.H.; Hsu, Y.H.; Lin, L.; Lin, Y.T.; López-Cobá, C.; Su, Y.C.; Accetta, K.; Beaton, R.L.; et al. The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. *Astrophys. J. Suppl. Ser.* **2022**, *259*, 35. [CrossRef]