Advanced Engineering Technology and Application
An International Journal

http://dx.doi.org/10.18576/aeta/130114

Machine Learning Approaches for Galaxy Categorization

Mohamed Ben Hassen[1*] *and Kaies Debbabi*[1]

[1] Textile Engineering Laboratory, Monastir University, ISET Ksar Hellal, Tunisia.

Received: 27 Jul. 2024, Revised: 2 Aug. 2024, Accepted: 5 Aug. 2024.
Published online: 1 Sep. 2024.

Abstract: The main topic of this study is the categorization of galaxies using machine learning methods on the 'Stellar categorization - SDSS17' dataset. The collection consists of 100,000 observations obtained using the Sloan Digital Sky Survey (SDSS). Each observation is identified as a star, galaxy, or quasar by one class column and 17 feature columns. Three classification models are used in the study: Decision Tree, XGBoost, and K-Nearest Neighbors (KNN). Exploratory data analysis provides insights into the distribution and correlations of the data after preprocessing of the dataset, which includes converting categorical class values to numeric values and removing unnecessary features. The models are then trained on the data, and different metrics such as accuracy, F1 score, confusion matrix, and classification ratio were used to assess how well the models performed. Based on the accuracy score of 97.5%, XGBoost performs better than the other models, with KNN and Decision Tree coming in second and third, respectively. Results obtained This work offers insightful information on the categorization of galaxies by machine learning methods, with possible uses in data analysis and astronomical research.

Keywords: Machine learning; Classification; Astronomy.

1 Introduction

The fundamental building blocks of the universe are galaxies, which are enormous, huge groupings of stars, gas, dust, and dark matter, each of which is home to harbor a different tapestry of cosmic occurrences. [1] The cosmos displays a vast assortment of galaxy shapes, each of which holds clues to its birth and evolution. These forms range varied from the magnificent spirals (that swirl gently across space) to the elliptical giants, spherical and calm, and the irregular galaxies, chaotic and turbulent.[2]

Of Among the many kinds of galaxies, three main categories predominate in our knowledge: spiral, elliptical, and irregular galaxies ones. They are formed by arms that stretch out extend from a central bulge, they are the beneficiaries of rich stellar creation and active galactic nuclei processes. This kind is generally demonstrated by our own galaxy, the Milky Way, whose spiral arms are filled with nebulae and clusters of swirling newly formed swirling stars [3]. On the contrary, elliptical galaxies, which appear as ovoid or flattened spheres, mixed with older star populations of older stars cover their vast areas. When the universe was young, all such of these galaxies would have had a uniform appearance comprised of stars, gas, and dust without any visible features such as spiral arms [4]

Contrary to the regular beauty of spirals and ellipticals, expressed in harmony and perfection, irregular galaxies stand out for their confusion and messiness. Galaxies like this which result from strong starbursts or being disturbed by nearby galaxies, may not have a systematic dynamic. Since a long time ago the main line of the research in astronomy research has been questioning to question the nature and age of the celestial bodies. Though thanks to the application of AI, new and more efficient methods of recording, classifying and analyzing objects are imminent the cosmological objects' registration, classification and analysis are imminent [5].

Artificial intelligence (AI) algorithms, especially machine learning ones, have uncovered the capability of processing astronomical data chunks to a much larger extent than ever before. This algorithm may discover that there is the connection between the nature of different various kinds of galaxies with such aspects as the spectra, statistical measurements and other observational data to that can the connection between the nature of different various of galaxies with such aspects as the spectra, statistical measurements, and other observational data to identify tiny patterns and signatures. The goal objective of our study is to bring this aspect of the area called "Artificial Intelligence - Astronomy interface" into focus by studying 'Stellar Classification - SDSS17' dataset (a huge database including stars, galaxies and quasars) with the help of using machine learning algorithms. This database includes data about the most important astronomical objects that we know of so far, including stars, galaxies and quasars. That is the result of findings from Sloan Digital Sky Survey (SDSS).

*Corresponding author e-mail: benrayen@yahoo.fr

A full machine learning approach was employed in the current study, encompassing data preprocessing, exploratory data analysis, model training, and performance evaluation. Our goal is to find out how well three popular models of galaxy classification do in precisely recognizing and classifying galaxies in the dataset: In this paper, KNN (K-Nearest Neighbors), XGBoost, and Decision Tree were used as the machine learning algorithms.

2 Related works

In a recent study, a machine learning method using photometric data was applied to classify different astronomical objects. Observed sources were selected from SDSS/DR17 and SDSS/ALLWISE databases [6]. This work studies the application of machine learning methods toward the classification of different astronomical objects ranging from stars to quasars and emission line galaxies (ELGs). This is done using carried out by making use of the photometric data. This work implies a two-stage method, which includes observed sources from the SDSS/DR17 sample, as well as a selection of sources from the SDSS/ALLWISE catalogue, considering the spectroscopically verified objects from SDSS/DR17. In order to come out with the best performance we would utilize the three-class, four-class, and seven-class classification tests carried out via the Random Forest (RF) technique during the first phase were utilized. Different from the encouraging result of the first phase, k-nearest neighbors (KNN), RF, XGBoost (XGB) and voting are learned in second phase which they include among others artificial neural networks (ANN). The optical and infrared characters should be utilized as very strong evidence as each solid proof in that every classifier has achieved the had best results being achieved. Surprisingly, the obtained accuracy achieved under the three-class model is identical for the RF and XGB algorithms, resulting in 98.93% of the F1 values on average.

The research showed the feasibility of employing automated algorithms to perform highly accurate do very precise configurations of the photometric data. This significantly facilitates studying the highly advanced topic of photometric categorization for collection of valuable insights which help to identify the subjects for further observations.

Here the critical component in grasping galactic evolution [nut] is the high-speed way of categorizing galaxies based on their morphology. This investigation focuses on this by using quantum computing. In addition to the effectiveness of SVM to classify galaxies, the new research method is quantum-enhanced SVM, and it is based on the large dimensions of the Hilbert space of the quantum. The method built to execute a quantum circuit which is too complex for classical simulation, computes kernel matrix on quantum simulator. These results support the idea that respective performances of traditional and quantum echo-sounding SVM algorithms on distinguishing between training sets of elliptical and spiral galaxies are practically the same measuring by ROC AUC value (0.946±0.005 for both of them). This data demonstrates that both for small data and big data, the results of simulations and quantum ensemble SVM algorithm are similar between the quantum device and the classical SVM algorithm. A fundamental quantitative condition for the success of the algorithmic computation of galaxy classification is discussed at the end of the paper. All things considered, the framework implemented in this paper is one of the first cases of quantum machine learning in astronomy, where we can see the tremendous potential for development of the approach in the future.

This technique [8] differs from the commonly used cosmic galaxy method and is based on an innovative machine learning approach that can be used to study the cosmological parameters. In another research paper [8], the machine learning model it was fitted to data, which is made of parameters such as gas mass, stellar mass, for example, bolometric brightness, temperature, cluster radius and total star mass, to obtain the necessary data using artificial catalogs from the Machine Learning Hydrodynamics simulations. The score values of cosmological parameters, such as Omega_m, sigma_8, Omega_b, and h_0, are predicted successfully by model, with about 14%, 8%, 6%, and 3% uncertainties, respectively. This novel test therefore realizes a significant opportunity for obtaining a much tightened much-tightened cosmological constraint by profiting from the success of machine learning in non- parametrically delineating the functional relations between observables and model parameters. Later on, machine learning tools would be one of the tools of deliver more accuracy in cosmological constraints and investigation of the effects of baryonic feedback through multi-wavelength data from surveys such as LSST, CSST, Euclid, Roman and eROSITA.

Fang et all. this work [9] aimed at solving the tough problem of defining galaxy morphologies in the presence the datasets produced by the upcoming generation of observatories. The polar-adaptive transform is a proposed approach by authors which refers to a rotationally invariant supervised machine learning approach (SML). The reliability of such an approach makes it possible to carry out conduct the process during the revolution of images, which is a task justified by physics, but also difficult to implement using algorithms in everyday performances. The polar- coordinate transformation is shown to be adaptive, as it seems to manifest better performances of solving issues fast in SML approaches than in the standard data augmentation process. The authors applied rotationally-invariant SML to classify galaxies into five types namely – irregulars, late-type disks, early-type disks, and spheroids) using the UML method with its catalog of galaxies feature well-separated by hand classification. The report of the method having the ability to automate the whole full process of classifying galaxy morphology is evidenced by the attained classifications showing the similarity in patterns when compared to features of other galaxies.

In this article wean another study, researchers introduced the C-GaMe system, an information-based machine learning for the determination of to determine the actual residents of clusters, i.e. falling in orbits, these ones coming in and background (interloper) ones [10]. As is the matter of the Hermean Universe Machine fake catalog, which is a basis of N- body simulation Multi-Dark Planck 2, it serves as the training and testing system for the algorithm. The research is useful in obtaining physical characteristics of clusters which are more accurately predictable by probabilistic technique as apparition of deterministic technique. The authors introduced new unbiased estimators for cluster features and demonstrate demonstrated that, even if there are intruders, the business can reconstruct an orbiting and falling galaxy with a high level of precision from a simulation of the entire phase space. The taught model's executable on the manifold simulation verifies their resiliency. The research in progress also tests the effect of including adjacent parameters such as star formation rate and galaxy halo mass divided by cluster halo mass among the cluster galaxy candidate's attributes to improve the precision of classification. Future directions for quenching knowledge about galaxies and cosmology of clusters are included in this method, as well.

Ndung'u this paper [11], including radio galaxy morphological classification, is thus contributed to a detailed review on the use of machine learning approaches in handling high amounts of data obtained from the most advanced future radio observatories such as the Square Kilometer Array (SKA). The paper illuminates the role of machine learning in the replacement of traditional statistics and in pointing out rare and uncommon astronomical incidents in the vast datasets for the cosmologists to investigate. They take a birdâ€™s-eye view in the most recent research that taps into problems the field facing, projects that are yielding good results, new technologies or methods, and possible future avenues. The article emphasizes the paradigm shift and automation revolution that have been triggered by the emergence of machine learning in radio astronomy modification. However, it still provides strong arguments that the astronomy research can never be realistically carried out without an effort from many team members especially in a multidisciplinary field needing the creation of high standard annotated data sets for facilities such as LOFAR and MeerKAT. The same manner, Wang [12] suggested the research use of the methods of radio galaxy (with the hypothesis that supposition that is chiefly based on vast data sets which do not only count on labeling output, but also data cubes that are made of multi-channel data).

The research objective was to spend more research time in the category of emission –line galaxies (ELGs) for the same reason. This task focuses on the labeling of the unidentified spectra through a creation of a sizeable database that contains spectra and the use of machine learning that would learn from the database. Specifically, the authors employed different machine learning approaches, e.g. random forest (RF), K-nearest neighbor (KNN), support vector machine (SVM) or multi-layer perceptron (MLP). With the system they invented based on those methods, they categorized 49,000 emission line galaxies (ELGs) observed by means of the Big Area Multi Object Fiber Spectrosc The spectral flux feature around commonly found emission lines, also known as the area of wavelength, where the feature set is extracted, is used in the classification process. Among the four following studied methods, MLP method garners the best 92.31% precision rate, the stats reveal.

The research also indicates the robustness of the MLP classifier and produces it for the classification of emission line galaxies with data acquired from fresh LAMOST images. The approach of ASID-C algorithm (ASID-C) introduced in this work [13] is geared to provide a direct image classifier for stars and galaxies from Astro photographic images. The basis of the CNN (Convolutional neural network) is 32x32 pixel single filter band input for the ASID-L (ASID-Light) algorithm. ASID-C (ASID-C) classifies the survey sources done by the ASID-L (ASID- Light) algorithm using the CNN (Convolutional neural network) and extra positional info. The alteration of Platt scaling calibration for modifying output probabilities of the method is the technique that makes the predictions trustworthy and correct. The result showed that the ASID-C outperforms SourceExtractor, which is a similar algorithm, based on the images taken by MeerLICHT telescope and the morphological classification Deep explaining about the dark energy used in the Dark Energy Camera Legacy Survey (DECaLS). The study also showed the performance excellence, computational efficiency, and error propagation decrease of Test-Case Sample Detection – Coarse (ASID-C) analytics when compared to the tabular features extracted by SourceExtractor instance of XGBoost model. In these scenes with lesser signal to noise ratio and high traffic the process of filtering and denoising is even more complex. The work highlighted the potential of ASID-C to improve deep-sky astronomy and identify transient hosts.

3 Methodologies

Data Acquisition and Preprocessing:

Processing starts with downloading the Slan Digital Sky Survey - SDSS17 Data Set. The dataset includes 100,000 observations. Each observation is made up of 17 feature fields and one column where class value indicates whether this object is a star, a galaxy or a quasar. The first step to this end is to put the dataset in a Pandas Data Frame which is better to use for data handling.

To make sure data is good for the machine learning, we use a number of preprocessing steps earlier. This is when we carefully go through the data to ensure that each element is present and is not clashing with another data point. If this is the case, we would correct the data using either the method of imputation or data removal. For example, we also digitize the simple expressions which that have been written in categorical class values.

Exploratory Data Analysis:

Using exploratory data analysis (EDA), the data was prepared, and the distribution, variable attributes, and correlation were described. Yes, the subsequent step was the data visualization and pattern mapping utilizing statistical plots, histograms, and correlation matrices. EDA is a method that looks at a dataset's class distribution in order to spot any imbalances that could have an impact on how well a classification algorithm performs as a model. However, analysis of the features' effects on the purpose variable (class) was also necessary to properly improve categorization. Features that added value were added, and those that were not didn't were removed.

Model Selection and Training:

Following preprocessing and exploratory data analysis (EDA), we train classification models such as Decision Tree, XGBoost, and K-Nearest Neighbors (KNN). These models are selected based on galaxy categorization, performance, and adaptability.

Using the train_test_split function from scikit-learn, we divided the dataset into training and testing sets prior to training the model. By examining the models' performance on fictitious data, we may assess how well they generalize.

Accuracy, F1 score, confusion matrix, and classification report are used to assess each model's performance once it has been trained on the training set with its default hyperparameters.

Machine learning approaches:

In order to create prediction models for the categorization of asteroids into hazardous and non- hazardous categories, we use a variety machine learning methods in this work. The pre-processed preprocessed dataset is used to train each method, and grid search combined with cross-validation is used to maximize performance. The machine learning techniques listed below are applied:

Decision Tree:

Using straightforward decision rules, Decision Tree is a non-linear classification technique that divides the feature space into several regions. In order to optimize class purity within each subset, it recursively divides the data into subgroups based on the most informative attributes. In this work, we build a decision tree classifier and use grid search to adjust hyperparameters such maximum depth, minimum samples per leaf, criterion for splitting (gini or entropy), and random state. We can understand the categorization criteria by looking at the decision tree that results, which gives us a clear visual representation of the decision-making process. The calculated feature importances, which illustrate the relative contributions of each feature one to the classification decision, are also used to examine feature importance.

XGBoost:

Gradient boosting is used by the ensemble learning method XGBoosting to create a group of ineffective learners (decision trees). It creates a sequence of decision trees one after the other, fixing the mistakes of the previous trees to produce a powerful learner with excellent predictive capability. In this work, we use grid search to adjust hyperparameters including random state, subsample ratio, column subsample ratio, learning rate, maximum depth, and regularization parameters (alpha and lambda). Additionally, XGBoosting has built-in feature significance metrics, which we display to determine which characteristics have the most influence on the classification process using bar plots and the plot_importanceplot importance function.

Following training, we use the testing dataset to assess the performance of the original models. We compare their accuracy and other metrics to ascertain how well the models categorize galaxies. This analysis directs further optimization and refinement by illuminating the advantages and disadvantages of each model.

We not only assess the models' performance but also investigate how model performance is affected by hyperparameter adjustment. We methodically adjust the hyperparameters of each model's hyperparameters using methods like such us grid search and randomized search to find the best configurations that optimize key metrics and classification accuracy.

K-nearest neighbor (KNN):

It is used to categorize issues categorization problems. The class of a data point' class is decided by the majority vote of its K closest neighbors in the feature space, according to the similarity or distance measure concept.

Here, cross-validation is applied to infer the value of K, known as the neighborhood size, which is especially important as it must be tuned properly to prevent obscuring the noise and over smoothing. FNN is a good method of learning with which it can even handle a 'nonlinear decision boundary' and 'non-linearly separable data' without training the data. In addition to that, the k-NN algorithm needs to take care of distance to all the data points, which is computationally expensive with a large data set. Additionally, the type of distance metric used, and the value of "K" can greatly affect the way network functions. And although KNN is realistic and effective on a variety of classifications and is being really often taken as a benchmark in comparison to more sophisticated algorithms, it has a couple of disadvantages that limit their algorithm's performance.

Evaluation metrics:

Model's the performance of the model was evaluated by multiple several measures at the end of when the training was over. Accuracy and loss were the two main indicators measured which helps to understand the generalization performance and the predictive power of the model as implied. In order to show the model's classification accuracy more vividly and find out where there may be a misclassification, a confusion matrix has been made. The accuracy, recall, and F1-scores were computed calculated for each category to determine the model's performance with different classes. Finally, reports were generated for making a detailed classification of the model performance depending on pros and cons and multiple evaluation metrics.

Classification metrics:

Accuracy: The accuracy is calculated by the percentage of correctly recognized samples out of relative to the total number of samples. It provides a tentative judgment of the fact that the model is competent in generating an accurate forecast.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

where:

TP: True Positives

TN: True Negatives

FP: False Positives

FN: False Negatives

Confusion Matrix: Giving the number of true positive, true negative, false positive, and false negative, the confusion

matrix brings the model's prediction analysis to a score. We can garner a set of other performance metrics, namely, the recall, precision, and F1-score, from the confusion matrix.

Precision: The percentage of true positives indicated among all predictions that are positive is measured by precision, which is also known as the positive predictive value. This illustrates how that algorithm can distinguish between true and false positives.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

where:

TP: True Positives

FP: False Positives

Recall, also known as sensitivity, is a metric that quantifies the proportion of accurate positive predictions out of all the actual positive samples in a given dataset. It quantifies the model's ability to accurately identify exemplary instances.

The F1-Score is a metric that quantifies the performance of a model by calculating the harmonic mean of precision and recall. It is suitable for datasets with skewed distributions as it considers both false positives and false negatives.

F1 score=2×(Precison ×Recall)/(Precision+Recall).

4 Results

Different factors are taken into account to evaluate categorization algorithm's, dividing quasars, galaxies and stars, accuracy, precision, recall and F1-score. These results highlight the evidence of different models for classification and asteroids models that guide us on effectiveness of each model is.

Decision Tree:

On the testing dataset, the Decision Tree model performed well with 96% accuracy. The categorization report indicates high accuracy, recall, and F1-score for each class. Precision, recall, and F1-score for class 2 (galaxy) were 100%, 100%, and 100%, respectively. Class 0 (quasar) had 97%, 97%, and 97% accuracy, recall, and F1-score values, whereas class 1 (star) had 91%, 91%, and 91%, respectively.

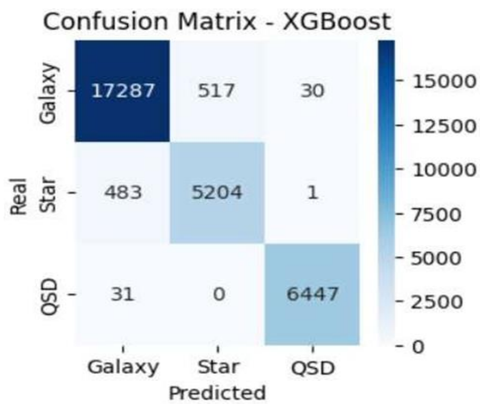
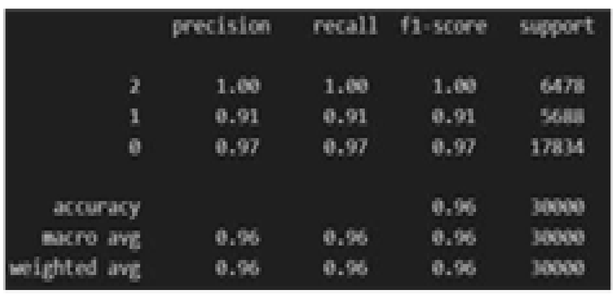


Fig. 1: Confusion matrix of XGBoost..



	precision	recall	f1-score	support
2	1.00	1.00	1.00	6478
1	0.91	0.91	0.91	5688
0	0.97	0.97	0.97	17834
accuracy			0.96	30000
macro avg	0.96	0.96	0.96	30000
weighted avg	0.96	0.96	0.96	30000

Fig. 2: Classification report of Decision tree.

XGBoost:

On the test dataset, the XGBoost model performed better than the others, producing an accuracy of 98%. All classes have good accuracy, recall, and F1-score values, according to the classification report. Precision, recall, and F1-score for class 2 (galaxy) were 99%, 99%, and 99%, respectively. Class 0 (quasar) had values of 98%, 98%, and 98%, respectively, whereas class 1 (star) had values of 96%, 93%, and 95% for accuracy, recall, and F1-score.

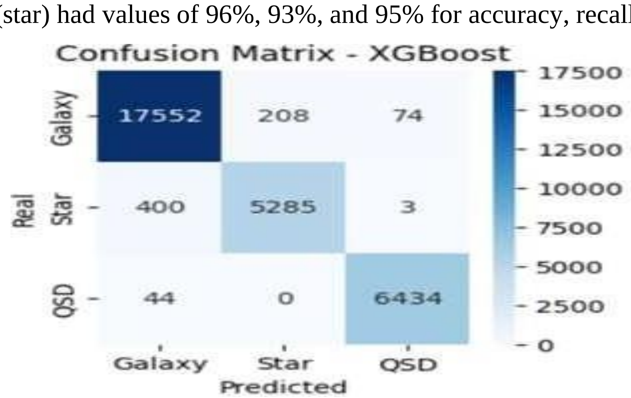


Fig. 3: Confusion matrix of XGBoost.

	precision	recall	f1-score	support
2	0.99	0.99	0.99	6478
1	0.96	0.93	0.95	5688
0	0.98	0.98	0.98	17834
accuracy			0.98	30000
macro avg	0.98	0.97	0.97	30000
weighted avg	0.98	0.98	0.98	30000

Fig. 4: Classification report of XGBoost.

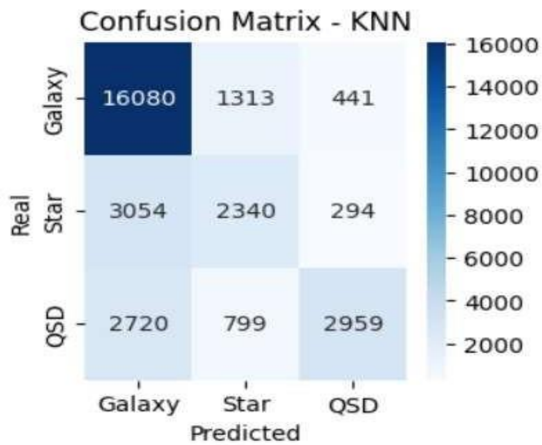


Fig. 5: Confusion matrix of KNN.

	precision	recall	f1-score	support
2	0.83	0.44	0.58	6478
1	0.52	0.41	0.46	5688
0	0.73	0.90	0.81	17834
accuracy			0.71	30000
macro avg	0.69	0.59	0.61	30000
weighted avg	0.71	0.71	0.69	30000

Fig. 6: Classification report of KNN.

K-nearest neighbor (KNN):

On the testing dataset, the accuracy of the KNN model was 71.26%. For every class, the classification report shows different values for accuracy, recall, and F1-score. The accuracy, recall, and F1-score for class 2 (galaxy) were 83%, 44%, and 58%, respectively. Class 0 (quasar) showed 73%, 90%, and 81% of accuracy, recall, and F1-score values, while class 1 (star) showed 52%, 41%, and 46% of these values. The accuracy of the model varied from 64.21% to 71.26% by changing the value of K from 1 to 24, with K=15 achieving the maximum accuracy.

5 Conclusions

This work demonstrated the ability of the machine learning algorithm XG Boost to categorize galaxies based on data from the Sloan Digital Sky Survey. XGBoost does better in preprocessing, training, and assessment than KNN and DT. Based on dataset testing, XGBoost achieved 98% accuracy. In studies of galaxy categorization, XGBost demonstrated enhanced robustness and durability. There were F1-score, recall, and accuracy numbers for every class. By accurately classifying and disclosing the attributes and classifications of the vast array of visible galaxies, AI has the potential to advance our understanding of the universe. It is anticipated that AI in astronomy would result in more sophisticated and in-depth understandings, igniting new study and unlocking more cosmic secrets. With 96% accuracy, the Decision Tree model demonstrated good performance on the testing dataset. Strong accuracy, recall, and F1-score are displayed in the categorization report for all classes.

Acknowledgment

We extend our heartfelt appreciation to the dedicated scientists and researchers whose contributions have been instrumental in the development of this track. Their pioneering work and invaluable insights have laid the foundation upon which our research builds. We are grateful for their innovative ideas, groundbreaking discoveries, and tireless efforts in advancing the field. Their expertise and commitment to excellence have inspired and guided our own endeavors, shaping the trajectory of this project. We deeply admire and acknowledge their significant contributions, which have enriched our understanding and propelled the progress of this track.

References

[1] Pilyugin, L. S. and Tautvaišienė, G., "Two sequences of spiral galaxies with different shapes of the metallicity gradients," A&A, vol. 682, p. A41, 2024, doi: 10.1051/0004-6361/202347032.

[2] P. Ghosh and S. Chakraborty, "Spectral Classification of Quasar Subject to Redshift: A Statistical Study," Comput. Sci. & Math. Forum, vol. 7, no. 1, 2023, doi: 10.3390/IOCMA2023-14418.

[3] J. Domínguez-Gómez et al., "Galaxies in voids assemble their stars slowly," Nature, vol. 619, no. 7969, pp. 269–271, 2023, doi: 10.1038/s41586-023-06109-1.

[4] A.A. Belinski et al., "Selected Active Galactic Nuclei from SRG/eROSITA Survey: Optical and IR Observations in 2021 and 2022 with the 2.5-m Telescope at the Caucasian Mountain Observatory of SAI MSU," Astrophys. Bull., vol. 78, no. 3, pp. 283–292, 2023, doi: 10.1134/S1990341323700074.

[5] W. Xiao-Qing and Y. Jin-Meng, "Classification of star/galaxy/QSO and star spectral types from LAMOST data release 5 with machine learning approaches," Chinese J. Phys., vol. 69, pp. 303–311, 2021, doi: https://doi.org/10.1016/j.cjph.2020.03.008.

[6] J. Z. Fatemeh Zahra Zeraatgari, Fatemeh Hafezianzadeh, Yanxia Zhang, Liquan Mei, Ashraf Ayubinia, Amin Mosallanezhad, "Machine learning-based photometric classification of galaxies, quasars, emission-line galaxies, and stars," Mon. Not. R. Astron. Soc. Vol. 527, Issue 3, vol. 527, no. 3, pp. 4677–4689, 2024, [Online]. Available: https://doi.org/10.1093/mnras/stad3436

[7] M. H. Hassanshahi, M. Jastrzebski, S. Malik, and O. Lahav, "A quantum-enhanced support vector machine for galaxy classification." 2023.

[8] L. Qiu et al., "Cosmology with Galaxy Cluster Properties using Machine Learning." 2023.

[9] G. Fang et al., "Automatic Classification of Galaxy Morphology: A Rotationally-invariant Supervised Machine-learning Method Based on the Unsupervised Machine-learning Data Set," Astron. J., vol. 165, no. 2, p. 35, Jan. 2023, doi: 10.3847/1538-3881/aca1a6.

[10] D. Farid, H. Aung, D. Nagai, A. Farahi, and E. Rozo, "C 2-GaMe: Classification of cluster galaxy membership with machine learning," Astron. Comput., vol. 45, p. 100743, 2023, doi: https://doi.org/10.1016/j.ascom.2023.100743.

[11] S. Ndung'u, T. Grobler, S. J. Wijnholds, D. Karastoyanova, and G. Azzopardi, "Advances on the morphological classification of radio galaxies: A review," New Astron. Rev., vol. 97, p. 101685, 2023, doi: https://doi.org/10.1016/j.newar.2023.101685.

[12] L.-L. Wang et al., "Spectral classification of LAMOST emission line galaxies based on machine learning methods," New Astron., vol. 99, p. 101965, 2023, doi: https://doi.org/10.1016/j.newast.2022.101965.

[13] Stoppa, F. et al., "AutoSourceID-Classifier - Star-galaxy classification using a convolutional neural network with spatial information," A&A, vol. 680, p. A109, 2023, doi: 10.1051/0004-6361/202347576.