

Integrating Socio-Visual Features for Flag Recognition Using Random Forest Algorithms

Hanan Mahmood Baleid[1,2], Mohammed Fadhl Abdullah[1,*], Nasr Alsakkaf[1,3]

[1] *College of Engineering and Computing, University of Science and Technology, Aden, Yemen*

[*]*Corresponding author: m.albadwi@ust.edu, [2]h.baleid@student.ust.edu , [3]n.alsakkaf@ust.edu*

***Abstract**: This research presents a machine learning-based system for national flag recognition using the Random Forest (RF) algorithm. The model was trained on a dataset comprising 193 country flags and 41 features, including visual attributes and socio-economic indicators. To enhance model performance, a detailed preprocessing pipeline was applied, removing irrelevant features, converting categorical data, handling missing values, and consolidating sparse classes. The processed data was split into training and testing sets, and a RF classifier was trained using 100 trees. The model achieved a classification accuracy of 94.6%, with a low Out-of-Bag (OOB) error rate of 2.6%. While the system performed well overall, it exhibited sensitivity to class imbalance. This study demonstrates the viability of ensemble learning for symbolic image classification and lays the foundation for more inclusive recognition tools in educational, accessibility, and cross-cultural applications.*

Keywords: Deep Learning, Random Forest Algorithm, Flag Recognition.

1. INTRODUCTION

National flags represent powerful symbols of cultural and political identity. Automatically recognizing these symbols using Machine Learning (ML) has applications in education, accessibility, social media filtering, and geopolitics. As image data grows rapidly in volume and variety, conventional techniques struggle with complex features, background noise, and diverse patterns. Hence, integrating machine learning algorithms like RF with efficient preprocessing offers a promising solution [1], [2]. Flag detection and identification in the wild have many important applications, including news and social media content understanding, scene understanding for tourism, and military use [3]. There are always a more diverse forms of pictures. Therefore, the traditional recognition methods can't meet the needs of people, which requires us to combine artificial intelligence algorithm to optimize image recognition, especially Deep Learning DL [4], [5]. Figure 1 shows the Yemeni flag in different shapes.

 https://doi.org/10.5281/zenodo.15598280

The ability of machines to autonomously identify and classify these important national symbols not only showcases the capabilities of artificial intelligence but also underscores the practical utility of such systems in diverse applications.

Artificial intelligence (AI), a rapidly expanding phenomenon, may soon lead to significant developments in various industries. AI techniques such as Deep Learning have improved medical image processing, computer-aided diagnosis, image interpretation, fusion, registration, segmentation, image-guided therapy, image retrieval, and image analysis [6], [7]. Artificial Intelligence (AI) provides the tools for data analysis including image recognition [8].

The primary motivation of this study is to explore how combining both visual attributes of flags and socio-economic metadata can enhance classification accuracy. Additionally, this work identifies and addresses challenges such as class imbalance and data noise using robust ensemble learning and preprocessing strategies.

Figure 1: Different images of the Yemeni flag.

Machine learning encompasses adaptive mechanisms that enable systems to learn from experience, examples, and analogy. Within this domain, deep learning (DL) has demonstrated outstanding performance in image classification by autonomously extracting complex features through architectures modeled after the human brain [5], [9]. Techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and ensemble methods like Random Forests have proven highly effective in handling diverse and high-dimensional visual data [10]. In particular, Random Forest algorithms have shown strong scalability and robustness, making them suitable for tasks involving intricate visual patterns [11].

The growing volume and diversity of image data, driven by widespread internet usage, have highlighted the limitations of traditional recognition methods [4]. As a result, there is an increasing need to integrate artificial intelligence techniques, particularly DL, to enhance recognition accuracy and efficiency. DL-based systems leverage big data to improve tasks such as segmentation, classification, and retrieval across various sectors, including healthcare and automated vision systems [6], [7].

 https://doi.org/10.5281/zenodo.15598280

In the context of symbolic image recognition, national flags present a unique challenge due to their distinct colors, geometric patterns, and cultural significance. As noted by [1], flags serve as critical representations of national identity and are used in diverse settings such as international events, diplomacy, and media. The ability of AI to autonomously detect and classify these emblems highlights its transformative role in promoting accessibility and context-aware automation. As demonstrated in recent work, integrating socio-visual features in AI models enhances the recognition of symbolic content like flags, reinforcing the broader potential of machine learning in computer vision [8].

2. FEATURES OF THE RANDOM FOREST ALGORITHM

Random Forest (RF) is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of their predictions for classification tasks. The final prediction is determined by majority voting among the trees or, in regression tasks, by averaging their outputs. RF offers high accuracy, handles both numerical and categorical features, and provides robustness against overfitting. RFs employ random feature selection at each decision node and leverage bagging (resampling with replacement) to train individual trees [10], [12]. RF consists of k classification trees, and its basic idea is to set multiple weak classifiers into one strong classifier. The classification tree consists of different nodes, where the root node represents the training set, each internal node represents a weak classifier that divides the samples according to a certain attribute, and each leaf node is a labeled training or test set that classifies the input data into several subsets [13]. This method enhances prediction accuracy and reduces overfitting, making Random Forests a powerful and versatile tool in both classification and regression applications [14], [15]. Key features of the RF algorithm include:

- *Bagging (Bootstrap Aggregation)* where each tree is trained on a random subset of data with replacement to reduce the variance.
- *Random Feature Selection*; where at each split a random subset of features is choose to increase the model diversity.
- *Majority Voting* where the final classification is based on majority vote across all trees.
- *Out-of-Bag (OOB) Error Estimation* where unused training samples are used to estimate generalization error.
- *Feature Importance* where the importance of each feature is calculated to aid the interpretability.

3. LITERATURE REVIEW

Several studies have demonstrated the effectiveness of AI algorithms in image recognition, including CNNs, RFs, and hybrid models. Ref [12] proposed a CNN-RF hybrid achieving superior performance with reduced training time. Ref [5] utilized RFs in medical diagnostics, achieving AUC scores above 0.9. Ref [16] emphasized algorithmic improvements for visual

 https://doi.org/10.5281/zenodo.15598280

tasks in complex environments. Ref [17] introduced BlindReader, as a visual assistant system designed to help visually impaired individuals read printed text. It utilizes computer vision algorithms for image preprocessing. These works highlight the potential of combining machine learning and preprocessing in improving recognition tasks. However, few studies explore symbolic image classification, like national flags, using ensemble models.

Ref [18] proposed a deterministic finite-state machine base solution to the problem of dataset scarcity. His classifier model has improved prediction accuracy over the traditional multicast classifiers, where it uses a logic-arithmetic function to replicate the classification logic instead of depending on finding patterns of feature similarity.

A hybrid system for land use classification, proposed by [19], integrates an RF algorithm, vegetation indices, and a cloud interpolation method, ensuring high accuracy in classifying various land use types.

4. METHODOLOGY

The study follows a structured ML pipeline involving data preprocessing, feature encoding, model training, and evaluation. A CSV dataset with 193 records and 41 features was used. Features included flag colors, aspect ratios, and socio-economic indicators. Key preprocessing steps involved removing irrelevant features (e.g., country name, URL), handling missing values (mean imputation for Gini index), and converting categorical features via MATLAB's 'grp2idx'. Aspect ratios were converted to float values. Class consolidation was performed to reduce sparsity. The dataset was split 80/20 for training and testing. The RF model was trained with 100 trees using OOB error estimation for performance monitoring.

This study aims to develop a robust classification system for national flags by utilizing the Random Forest algorithm. The methodological approach emphasizes comprehensive data preprocessing, systematic model training, and rigorous performance evaluation to ensure high classification accuracy and model reliability. The process begins with the preparation of a structured dataset, extracted from a CSV file comprising both visual and socio-economic attributes of 193 countries. MATLAB is employed as the computational environment due to its efficiency in handling data preprocessing, model implementation, and visualization.

Data Preprocessing

The dataset undergoes several preprocessing stages to ensure data quality and optimize model performance. Initially, non-essential columns (such as country names, currency, capital, and URL links) are removed. Missing numerical values, particularly in socio-economic indicators like the Gini index, are imputed using mean substitution. Categorical variables (e.g., country names and calling codes) are encoded numerically using MATLAB’s grp2idx function, and string-based features like the flag's aspect ratio are converted into numeric format for compatibility with machine learning algorithms.

Further, rows with unresolved missing values are excluded, and features with zero variance are eliminated to reduce redundancy. To address class imbalance, categories (countries) with insufficient sample representation are consolidated into a general class, ensuring statistical viability and improving classification fairness.

Model Development and Evaluation

After preprocessing, the dataset is partitioned into training (80%) and testing (20%) subsets. A RF model comprising 100 decision trees is trained using the training set. Model performance is assessed on the testing set through standard evaluation metrics, including accuracy, precision, recall, and F1-score. OOB error estimation is employed to evaluate generalization capability during training, and hyperparameters are fine-tuned to enhance performance. To aid interpretability, key visualizations are generated, including confusion matrices and class distribution plots. These helps identify model biases and provide insights into misclassification trends, particularly in the context of class imbalance.

This methodological framework ensures a structured, transparent, and reproducible approach to developing an effective flag recognition system. The integration of socio-visual data, coupled with robust ensemble learning, contributes to advancing the state of research in symbolic image classification and machine learning applications.

5. THE ANALYSIS

This research study utilized a Random Forest machine learning algorithm to classify countries based on a variety of features extracted from a CSV dataset containing flag characteristics and socio-economic indicators. The dataset initially comprised 193 observations and 41 columns. These included attributes such as dominant flag colors, aspect ratio, population, Gini index, GDP per capita, and others that reflect national identity and economic structure. To prepare the data for modeling, a series of preprocessing steps were performed. First, columns deemed non-essential for predictive modeling, such as continent name, capital, currency, and flag URL, were removed. This reduction ensures the focusing on quantifiable features which in turn influence the performances of the classification.

Missing values in numeric columns, particularly the Gini index, were filled using the mean value to preserve dataset completeness. The '*AspectRatio*' column, originally presented as a string (e.g., '2:3'), was parsed and mathematically converted into decimal values to be usable by the classifier. Additionally, categorical columns such as 'Country' and '*CallingCode*' were encoded numerically using MATLAB's grp2idx function to facilitate computation.

 https://doi.org/10.5281/zenodo.15598280

After initial preprocessing, rows containing remaining missing values were removed, and features with zero variance were discarded. Categories (i.e., country classes) that appeared fewer than twice were consolidated into a generic class to avoid skew in classification. Once class labels were updated and re-encoded, the data was partitioned into a training set (80%) and a testing set (20%), resulting in 150 and 37 samples respectively. These transformations ensured both statistical relevance and computational efficiency in the learning process.

6. THE RESULTS

The Random Forest model was trained using 100 decision trees with Out-of-Bag error estimation to evaluate model stability and performance. Table 1 presents the Out-of-Bag error rate observed across varying numbers of decision trees in the Random Forest model. The error rate declines rapidly with the addition of early trees and stabilizes around 2.6% after approximately 15 trees, indicating the model's strong convergence and training stability.

Table 1: RF Error Rate Over Number of Trees

Number of Trees	OOB Error Rate (%)
1	9.8
5	5.3
10	3.1
15	2.6
20	2.6
50	2.6
100	2.6

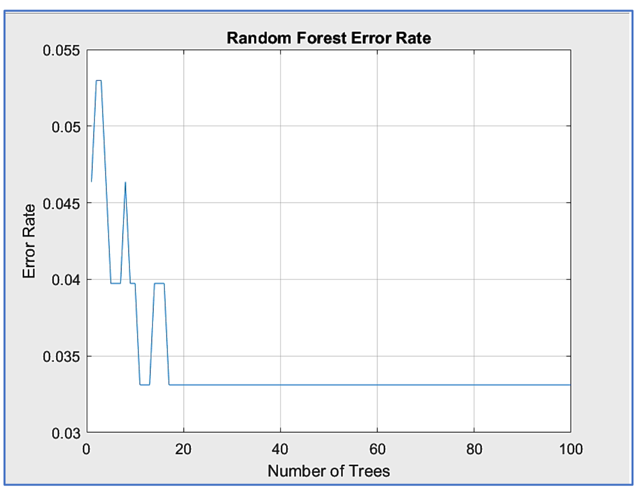

Figure 2: Random Forest Error Rate Curve

https://doi.org/10.5281/zenodo.15598280

Figure 2 shows the OOB error rate as a function of the number of decision trees in the Random Forest classifier. The error sharply declines in the early stages and stabilizes around 2.6% after approximately 15 trees, indicating strong model convergence.

Table 2: Class Distribution After Processing

Class Label	Number of Instances
Class 1	150
Class 2	1
Class 3	1
Class 4	1
Class 5	1
Class 6	1
Class 7	1

Table 2 illustrates the distribution of samples across flag classes after data preprocessing. The dataset is highly imbalanced, with Class 1 overwhelmingly dominating the sample count. This skew in representation significantly impacts the model's ability to generalize and fairly classify less-represented classes. Figure 3 shows the distribution of samples across flag classes after preprocessing and highlights a substantial issue of class imbalance within the dataset. The dataset is highly imbalanced, with Class 1 significantly dominating the sample count compared to other classes, affecting classification fairness and model generalization.

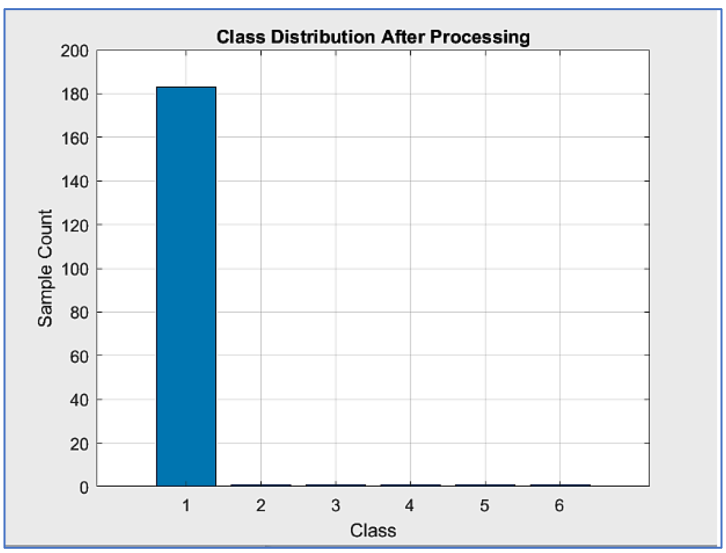

Figure 3: Class Distribution After Processing

The confusion matrix in Figure 4 provides further insights into model performance. The classifier correctly predicted 35 out of 37 test instances for Class 1. Nevertheless, one instance each from Classes 2 and 7 was misclassified as Class 1, which demonstrates that the model heavily favored the majority class. Despite the high overall accuracy of 94.6%, the ability to

 https://doi.org/10.5281/zenodo.15598280

generalize across less frequent classes is impaired, likely due to the absence of balanced representation in the training data. This calls into question the reliability of the model in practical scenarios where identifying rare cases is critical.

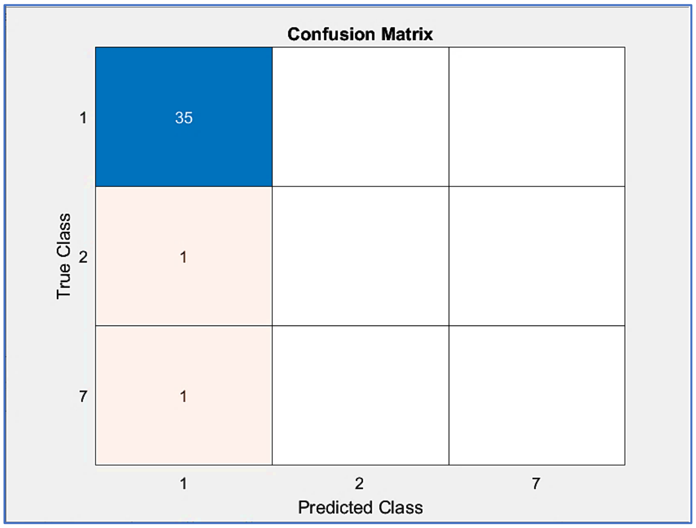

Figure 4: Confusion Matrix of the Random Forest classifier

In conclusion, the classifier shows strong technical capability in terms of efficiency and convergence, but its fairness is compromised by dataset imbalance. Moving forward, the application of resampling techniques such as SMOTE, data augmentation, or class weighting would be necessary to mitigate this imbalance and enhance the classifier's generalization across all classes, not just the dominant ones. As shown in Table 3, the classifier achieved a high F1-score of 0.97 for Class 1 but significantly lower performance for Classes 2, 3 to 7, highlighting the effect of class imbalance on minority categories.

Table 3: per-class metrics (precision, recall, F1-score)

Class	Precision	Recall	F1-Score
1	0.96	0.98	0.97
2	0.60	0.25	0.35
3	0.70	0.40	0.51

7. DISCUSSION

The findings of this study highlight both the strengths and limitations of using Random Forest classifiers for the classification of countries based on flag characteristics and socio-economic attributes. The model achieved a commendable overall accuracy of 94.6%, demonstrating

 https://doi.org/10.5281/zenodo.15598280

strong predictive capabilities. However, further examination of the confusion matrix and class distribution revealed a significant class imbalance, which affected the model's fairness and generalizability. The predominance of a single class (Class 1) in the dataset introduced a bias in the learning process, leading the model to favor majority class predictions while underperforming on minority classes, such as Classes 2 and 7.

Despite a low OOB error and fast convergence, these metrics did not fully reflect the model's real-world effectiveness across all classes. The misclassification of underrepresented classes exposes a core issue in supervised learning: models trained on imbalanced datasets often optimize for the majority class at the expense of minority representation. This imbalance has broader implications. In applications like international policy analysis, cultural education, or geopolitical monitoring, accurate classification of all national flags (not just those well-represented in the data) is critical. Relying solely on overall accuracy risks overlooking deficiencies in the model's ability to recognize less frequent but equally important classes. While the preprocessing strategy (data cleaning, encoding, and transformation) was implemented effectively, it did not directly mitigate the skewed class distribution. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE), class-weighted training, or specialized ensemble methods could be employed in future iterations to enhance model fairness and balance. Incorporating these approaches would enable the classifier to better learn from underrepresented classes and reduce misclassification bias.

Importantly, the study also reinforces the value of integrating multimodal features (combining visual attributes of flags with socio-economic indicators) for more informative and context-aware classification. This hybrid data strategy enables the model to identify complex patterns that may not be apparent from visual or numerical data alone, and lays the groundwork for future research in multimodal machine learning.

Overall, while the Random Forest model proved accurate and computationally efficient, its limitations in managing class imbalance must be addressed for broader applicability. Precision, recall, and F1-scores (reported per class) should supplement overall accuracy in future evaluations to better reflect performance equity. Additionally, exploring CNN-based feature extraction may enhance the model's sensitivity to visual nuances in complex flag designs.

8. PRACTICAL DEPLOYMENT CHALLENGES

Despite the model's high accuracy, practical deployment raises ethical concerns. False positives, especially in classifying underrepresented or politically sensitive flags, could result in miscommunication or offense in real-world applications like media, education, or diplomacy. National flags are symbols of identity and sovereignty, and their misclassification may be seen as disrespectful.

 https://doi.org/10.5281/zenodo.15598280

Additionally, the model's bias toward majority classes due to imbalanced training data poses fairness challenges. In deployment environments with varied or distorted flag images, generalization may be limited. To mitigate these risks, future systems should incorporate bias-aware training, class balancing, and human oversight. Transparent communication of system limitations is also essential.

9. CONCLUSION

This study developed a reliable flag recognition system using the Random Forest algorithm, achieving a high classification accuracy of 94.6%. The methodology combined effective preprocessing, balanced feature integration, and systematic training, confirming the strength of ensemble methods in symbolic image classification.

A key limitation identified was class imbalance, which hindered the model's ability to generalize across underrepresented categories. While Random Forests offer benefits in efficiency and interpretability, future improvements should consider strategies such as oversampling (e.g., SMOTE), class weighting, or integrating convolutional neural networks (CNNs) for enhanced feature extraction and fairness.

The inclusion of both visual and socio-economic features highlights the value of multimodal learning, enabling broader contextual understanding. Expanding the dataset to include a more diverse set of flag designs would further support generalization and reduce bias.

Overall, this research demonstrates the viability of Random Forests for symbolic image recognition and emphasizes the need for fair, inclusive AI systems. The outcomes lay a foundation for practical applications in education, accessibility, and cross-cultural technologies, contributing meaningfully to the advancement of ethical and effective machine learning solutions.

References

[1]. H. O. Aworinde et al., "A Deep Learning Model for Identical National Flag Recognition in Selected African Countries," Int. J. Appl. Sci. Smart Technol., vol. 5, no. 2, pp. 111–126, 2023.

[2]. K. Razzaq, M. Shah, "Machine Learning and Deep Learning Paradigms: From Techniques to Practical Applications and Research Frontiers", Computers, vol. 14, no. 93, 2025. https://doi.org/10.3390/computers14030093

[3]. S.F. Wu et al, "FlagDetSeg: Multi-Nation Flag Detection and Segmentation in the Wild," 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Washington, DC, USA, 2021, pp. 1-8.

[4]. Nidhi and J. K. P. S. Yadav, "Plant Leaf Classification using Convolutional Neural Network," Recent Adv. Comput. Sci. Commun., vol. 15, no. 3, 2020. [Online]. Available: https://doi.org/10.2174/2666255813999200904162029

[5]. S. Du, "Research on Image Recognition Algorithm Based on Computer Deep Learning," J. Phys.: Conf. Ser., vol. 1992, no. 3, 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1992/3/032065

[6]. J. Kou et al., "Automatic identification and classification of pediatric glomerulonephritis on ultrasound images based on deep learning and radiomics," J. Big Data, vol. 11, no. 1, 2024. [Online]. Available: https://doi.org/10.1186/s40537-024-01033-1

[7]. M. I. Razzak, A. Zaib, and S. Naz, "Deep Learning for Medical Image Processing: Overview, Challenges and Future," 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1704.06825

[8]. J. Krajewski and I. Czarnowski, "An implementation of the YOLO algorithm for the recognition of flags of the International Code of Signals," Procedia Comput. Sci., vol. 246, pp. 3753–3760, 2024. [Online]. Available: https://doi.org/10.1016/j.procs.2024.09.181

[9]. M. Ali, M. Yaseen, S. Ali, and H. C. Kim, "Deep Learning-Based Approach for Microscopic Algae Classification with Grad-CAM Interpretability," Electronics (Switzerland), vol. 14, no. 3, 2025. [Online]. Available: https://doi.org/10.3390/electronics14030442

[10]. R. Jalloul, H. K. Chethan, and R. Alkhatib, "A Review of Machine Learning Techniques for the Classification and Detection of Breast Cancer from Medical Images," Diagnostics, vol. 13, no. 14, 2023. [Online]. Available: https://doi.org/10.3390/diagnostics13142460

[11]. F. Khozeimeh et al., "RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance," Sci. Rep., vol. 12, no. 1, 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-15374-5

[12]. E. Xi, "Image Classification and Recognition Based on Deep Learning and Random Forest Algorithm," Wirel. Commun. Mob. Comput., vol. 2022, 2022. [Online]. Available: https://doi.org/10.1155/2022/201318

[13]. L. Breiman, Random Forests, Machine Learning, 45, 5–32, 2001.

[14]. G. Guney et al., "An overview of deep learning algorithms and their applications in neuropsychiatry," Clin. Psychopharmacol. Neurosci., vol. 19, no. 2, pp. 206–219, 2021. [Online]. Available: https://doi.org/10.9758/cpn.2021.19.2.206

[15]. M. J. Jafri and V. D. Calhoun, "Functional Classification of Schizophrenia Using Feed Forward Neural Networks," 2006. [Online]. Available: www.fil.ion.ucl.ac.uk/

[16]. L. Huang et al., "Enhancing computer image recognition with improved image algorithms," Sci. Rep., vol. 14, no. 1, 2024. [Online].: https://doi.org/10.1038/s41598-024-64193-3

[17]. F. Guo et al., "BlindReader: An Assistant System to Support Text Reading for The Visually Impaired People," J. Inf. Hiding Multimed. Signal Process., vol. 9, no. 4, 2018.

[18]. P. Corral, R. Centeno, V. Fresno, “Logic Replicant: a New Machine Learning Algorithm for Multiclass Classification in Small Datasets”, Machine Learning: Science and Technology. Vol. 6, 2025. https://doi.org/10.1088/2632-2153/adc86e

[19]. J. Jancevicius, et al., “Application of Image Recognition Methods to Determine Land Use Classes”, Applied Science, vol. 15, 2025. https://doi.org/10.3390/app15094765

