ORIGINAL RESEARCH **Open access**

Optimal Feature Engineering for Automated Red Rice Variety Classification Using Machine Learning

James Anderson[1], Maria Rossi[2*]

Abstract

Studies to date have largely focused on classifying white rice using readily accessible datasets. In contrast, the automatic categorization of red rice types remains an underexplored domain, with no publicly available datasets, thereby leaving a notable void in the field of agricultural image processing. This investigation examines red rice classification, a largely uncharted territory with few publicly shared datasets or targeted research on red rice variety recognition. The study classifies three specific red rice cultivars—Uma, KCP-1, and Jyothi—predominantly grown in Karnataka and Kerala using image processing and machine learning methodologies. Six ML classifiers were tested with seven unique groupings of features derived from size, shape, and texture traits to pinpoint the most distinguishing feature collection. Feature selection was performed using Recursive Feature Elimination and Backward Feature Elimination to improve model performance. Hyperparameter optimization was performed to maximize classification performance, and k-fold cross-validation, coupled with statistical significance testing, was used to assess generalization and confirm performance differences among models. The fusion of size, shape, and texture attributes yielded the highest mean accuracy across the classifiers, with K-Nearest Neighbours attaining 98.67% and Support Vector Machine attaining 97.34% when using the size and shape feature set. The outcomes underscore the importance of careful feature selection and parameter tuning for improving classification precision, supporting the development of automated sorting systems for red rice cultivars.

Keywords Image processing, Feature extraction, Feature selection, Red rice classification

*Correspondence:
Maria Rossi
maria.rossi@unimi.it
[1] Department of AI for Precision Crop Yield Prediction, Faculty of Agriculture, University of Manchester, Manchester, United Kingdom
[2] Department of Smart Irrigation and IoT Sensor Networks, Faculty of Agricultural Sciences, University of Milan, Milan, Italy

Introduction

Red rice cultivars are gaining international popularity for their remarkable nutritional properties and unique cooking characteristics. Across South India, red rice constitutes a dietary cornerstone that harmonizes with local eating patterns and cultural customs. Within the diverse array of varieties, Uma, KCP-1, and Jyothi stand out for their distinct flavor profiles and wellness attributes, aligning seamlessly with the modern shift toward more nutritious food choices. In particular, the Uma cultivar, raised through organic methods and ancestral farming techniques at Kerala Agricultural University, is highly regarded for its substantial iron and calcium concentrations. This strain is especially suitable for individuals with diabetes, as it has a lower glycaemic index than polished white alternatives. The time-honored procedures of parboiling followed by sun-drying, applied during Uma processing, significantly enhance its nutritional value while preserving its inherent characteristics. Sourced from a single farm, the grain maintains consistent quality and taste; it undergoes parboiling in a large vessel over a wood-fueled flame, a technique that helps preserve the majority of the bran's micronutrients, positioning it as a nourishing option in the daily diet [1, 2].

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

Turning to KCP-1, another distinguished cultivar, it is adapted to the coastal agricultural belts of Karnataka. It is recognized for its strong yield potential and exceptional straw quality, making it an economically sound alternative for cultivators. This variety reaches maturity in about 130–145 days and can yield 6-6.5 tons per hectare, a figure that substantially surpasses that of numerous competing varieties. The grains are lengthy and bold, suited not merely for lowland farming systems but also prized for their behavior during cooking. Beyond yield, the KCP-1 strain shows some tolerance to various paddy ailments and insect pests, including the gall midge, further enhancing its appeal to agronomists and growers seeking dependable cropping options. Similarly, Jyothi rice, also known as Vadi matta, is primarily cultivated in Palakkad and is acclaimed for its high dietary fiber content and dense nutritional profile. It takes the form of an unpolished, parboiled grain that preserves a considerable share of its native nourishment, solidifying its status as a wholesome foundation of South Indian gastronomy. The widespread embrace of this variety stems from its flexibility across a spectrum of traditional preparations, spanning routine meals to celebratory culinary creations, thereby illustrating the convergence of agrobiodiversity and nutritional imperatives [1, 3].

These red rice varieties not only signify notable agricultural breakthroughs but also serve as vital tools for addressing food security challenges. Their fortified nutrient profiles, together with their adaptability to indigenous agronomic settings, reveal the latent capacity of both heritage and upgraded cultivars to advance health-conscious diets while underpinning ecologically sound farming systems in South India. In this light, scrutinizing these varieties offers meaningful insights into the intersection of dietary wellness and cultivation practices, underscoring the imperative of safeguarding and championing native crop germplasm in the fight against malnourishment and lifestyle-related health conditions.

The proposed scheme for the automatic classification of red rice varieties faces a crucial deficiency in grain categorization, where automation has been overwhelmingly directed at white rice, bypassing red rice due to the absence of purpose-built datasets and scholarly inquiry. This work puts forward an original framework for differentiating Uma, KCP-1, and Jyothi cultivars, deploying an all-encompassing methodology that begins with the gathering of 300 kernel samples (100 per variety) sourced from the Zonal Agricultural & Horticultural Research Station, Brahmavar, and proceeds through uniform image capture utilizing a 45-megapixel sensor under d-65 illumination. The dataset construction phase included preprocessing steps (grayscale conversion, Otsu's thresholding, morphological processing), followed by the derivation of dimensional, form-related, and surface-texture descriptors. The feature integration process systematically examined every viable combination to determine the optimal descriptor ensemble, culminating in a peak accuracy of 98.67% with the K-Nearest Neighbours algorithm. The feature selection stage, a principal novelty of this work, used Recursive Feature Elimination (RFE) and Backward Feature Elimination (BFE) to remove superfluous descriptors, thereby streamlining model economy. Conventional machine learning architectures were deliberately chosen given the constrained dataset volume (100 instances per category) and the demand for computationally lightweight, readily interpretable algorithms suitable for deployment in agricultural contexts. Model performance was further refined through hyperparameter tuning, while adopting k-fold cross-validation (k = 5), statistical significance tests, and graphical diagnostics via confusion matrices and ROC curves ensured rigorous validation and extrapolative reliability. Altogether, this pipeline strengthens quality oversight and cultivar verification along the agricultural value chain, equipping stakeholders to respond to the growing demand for nutritionally dense red rice across South India.

Motivation

The impetus for this inquiry arises from the paucity of scholarly work on automated sorting of red rice cultivars and the concurrent lack of publicly released benchmark datasets. Although machine learning methodologies have been used to categorize white rice and other crops, prior investigations have entirely overlooked specialized red rice landraces such as Uma, Jyothi, and KCP-1. The present endeavor seeks to close this lacuna by engineering a resilient classification pipeline that can bolster the effectiveness of the logistical chain, commodity valuation, and quality enforcement. As consumer appetite for red rice grows due to its health-enhancing properties, precise identification techniques become indispensable for safeguarding product consistency and meeting market benchmarks. Furthermore, red rice cultivars exhibit idiosyncratic morphological and surface-textural signatures that mandate bespoke taxonomical strategies extending well beyond those formulated for white rice. The construction of this pioneering framework will not only augment the corpus of

© 2022 The Author(s). | Published by Agro Intelligence Research Press

academic understanding but also advance the field. Still, it will also yield actionable instruments enabling stakeholders to authenticate varieties and curtail fraudulent substitution in the marketplace.

Key contributions and novelty

• Confronts an unexplored void in the literature by devising an automated sorting pipeline specifically engineered for red rice cultivars, a domain left untouched until now.

• The lack of a publicly accessible repository of red rice images compelled the assembly of a purpose-built, high-resolution photographic collection for this work.

• Attribute computation followed by reduction was performed using Recursive Feature Elimination and Backward Feature Elimination to identify the subset of features with the greatest discriminative power.

• Systematic hyperparameter tuning was performed to optimize model settings and achieve peak classification performance.

• A side-by-side evaluation spanning six machine learning paradigms, including an ensemble strategy, was conducted to identify the algorithm best suited for distinguishing red rice types.

• Earlier efforts centered on white rice sorting relied on deep learning pipelines that either entail steep computational overhead, incorporate an inflated feature space replete with non-informative variables, or deliver suboptimal accuracy. In contrast, the present work delivers a resource-conscious alternative grounded in judicious feature pruning and model refinement, yielding markedly improved outcomes.

Related Works

Image processing coupled with machine learning has been fruitfully applied within the rice sector for kernel sorting, as shown by Arora *et al.* [4]. They constructed a pipeline using Logistic Regression, Decision Tree, Naïve Bayes, KNN, and SVM to differentiate among four grain types—Surti Kolam, Idli Rice, Long Grain Basmati, and Boiled Rice—aiming to increase industrial productivity through mechanized separation. Nagoda and Ranathunga [5] deployed a machine vision-based apparatus for rice quality grading powered by SVM. Their framework uses the Watershed Algorithm, Local Binary Pattern (LBP) surface descriptors, and colour-based cues to partition rice into whole kernels, broken fragments, and residual categories, achieving a segmentation fidelity of 96% and a classification rate of 88%. The contribution by Kiratiratanapruk *et al.* [6] describes an automated system for paddy seed sorting using machine vision. That system, built on InceptionResNetV2 and SVM, distinguishes 14 discrete rice cultivars by exploiting shape, colour, and surface attributes, achieving an accuracy of 95.15%. Probing the reach of deep learning for grain taxonomy, Aukkapinyo *et al.* [7] enlisted Mask R-CNN to locate and sort rice kernels. Their model improves on conventional pipelines with substantial gains in accuracy, underscoring deep learning's ability to handle grains with near-identical appearances. Krishna *et al.* [8] described an automated rice quality grading framework reliant on image processing and morphological procedures. Their scheme, which capitalizes on MATLAB for descriptor computation and SVMs along with Neural Networks for assignment, outperforms human-dependent inspection practices.

Carrying forward this wave of innovation, Ibrahim *et al.* [9] crafted a multiclass SVM-centric classification platform for rice kernels. Shape and colour descriptors were computed, substantially accelerating grading speed and reliability in the agro-tech sector while achieving 92.22 % accuracy. Ruslan *et al.* [10] proposed a scheme for detecting weedy rice that drew on a broad battery of 67 morphometric, chromatic, and textural variables. Their methodology, which canvassed numerous machine learning classifiers, achieved a top accuracy of 97.9% with Logistic Regression. A cascaded network philosophy was embraced by Singh *et al.* [11] for rice grain categorization. They fused a Backpropagation Neural Network with a fuzzy inference layer, boosting sorting precision via morphological descriptors and demonstrating resilience when tested

© 2022 The Author(s). | Published by Agro Intelligence Research Press

on UCI repository data. Cinar and Koklu [12] scrutinized a large-scale corpus comprising 75,000 rice grain images through a diverse suite of machine learning models. Their investigation underlined the decisive role of feature curation, with the Multilayer Perceptron model approaching flawless accuracy. In a comparative benchmarking of deep learning architectures, Farahnakian *et al.* [13] evaluated the performance of ResNet and MobileNet on a large rice-grain dataset. They remarked that EfficientNet not only achieved the highest accuracy but also highlighted the throughput advantages that deep learning confers on mechanized classification workflows.

Moving on to further studies, Nga *et al.* [14] refined feature subset selection by coupling Binary Particle Swarm Optimization with SVM, thereby improving classification rates across 17 Vietnamese rice accessions. Kuo *et al.* [15] adopted a sparse-representation-based categorization scheme that demonstrated both robustness and computational efficiency, making it especially well-suited for confirming rice authenticity from microscopic images. Exploiting Near-Infrared Spectroscopy alongside machine learning, Carneiro *et al.* [16] estimated milled rice quality; their models uncovered a strong linkage between physicochemical indices and grain grade, confirming the promise held by such techniques for non-invasive testing. The wide-ranging survey conducted by Ahmed *et al.* [17] charted the developmental trajectory of rice grain analytical tools, noting a pivot toward deep learning methods that have consistently returned superior accuracy in both sorting and disease spotting.

Srimulyani and Musdholifah [18] investigated the capacity of neural networks to classify rice accessions using geometric descriptors, achieving significant improvements in accuracy. Ksh examined the deployment of a Backpropagation Neural Network reinforced by wavelet decomposition. Singh *et al.* [19], whose results revealed an edge over standard classifiers attributable to the sturdiness of wavelet-driven feature sets. Aznan *et al.* [20] demonstrated the viability of smartphone-sourced images for commercial rice kernel sorting, highlighting the on-the-ground utility of an ANN combined with Bayesian Regularization for instantaneous quality appraisal. While reviewing ANN-based methodologies, Hamzah *et al.* [21] highlighted the promise of hybrid paradigms for improving both classification accuracy and quality grading of white rice grains. Sultana *et al.* [22] deliberated on the role of quantitative trait loci in rice grain quality, advocating molecular breeding strategies to refine rice quality parameters.

Ahad *et al.* [23] examined CNN architectures for the detection of rice pathologies, spotlighting the strength of ensemble configurations in real-time disease surveillance. The inquiry conducted by Tran-Thi-Kim *et al.* [24] underscored the edge of CNNs over conventional ML pipelines for rice variety differentiation, capitalizing on advanced image manipulation and data augmentation techniques. Patel and Sharaff [25] rolled out a deep learning-driven solution for simultaneous rice variety classification and yield forecasting, showcasing the relevance of semantic segmentation within this sphere. Kurade *et al.* [26] assembled an economical, automated rice quality inspection unit centred on a Raspberry Pi module, which delivered robust operation across a range of rice types. Deepika *et al.* [27] used Linear Discriminant Analysis in conjunction with digital imaging to grade rice kernel quality, achieving high accuracy in distinguishing high-yielding aromatic hybrid lines. Philip and Anita [28] unveiled an original sorting framework rooted in the Fourier Transform and morphological descriptors, achieving higher classification accuracy across an array of South Indian rice landraces.

Although prior investigations [4-6] have advanced rice sorting by leveraging image processing and machine learning, none of those efforts address red rice accessions such as Uma, KCP-1, and Jyothi, which have substantial dietary value and are witnessing escalating demand across South India. These accessions possess idiosyncratic morphological signatures—embracing colour depth, kernel contour, and surface patterning—that necessitate purpose-built taxonomic pipelines. A considerable portion of deep learning-centric studies [7, 13, 24] hinge on massive data repositories and computationally heavyweight frameworks, diminishing their suitability for resource-constrained settings where premium hardware and extensive labelled collections are unavailable. What is more, training deep architectures on sparse datasets habitually induces overfitting and compromises extrapolative capacity. Other contributions [9, 10, 12] use bloated descriptor suites that may include redundant or non-informative variables, thereby escalating processing overhead and eroding model intelligibility.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

By contrast, the approach put forward here zeroes in on red rice discrimination using parsimonious classical machine learning paradigms trained on a lean dataset (100 samples per category), achieving 98.67% accuracy via disciplined feature culling techniques, namely Recursive Feature Elimination and Backward Feature Elimination. This selective regimen ensures that only morphological descriptors with the highest information content are retained, resulting in faster training and inference cycles. Furthermore, the model blueprint lends itself readily to deployment on field-level hardware or handheld agricultural sensing devices. In contrast to earlier reports [8, 28] that omit rigorous cross-validation, the present inquiry employs k-fold cross-validation (k = 5) and statistical significance testing to ensure dependable generalizability. By centring on red rice, condensing feature ensembles, and privileging computational economy, the suggested pipeline constitutes a workable, high-performing remedy for automated red rice categorization that is at once scalable and applicable within authentic agricultural contexts.

Proposed Method

The experimental protocol adopted in this investigation, schematically depicted in **Figure 1**, commences with sample procurement and photographic capture, followed by dataset grooming to ensure optimal preprocessing. Three families of descriptors are harvested: size, shape, and texture. These attributes undergo fusion and are subsequently distilled via feature selection before being supplied to a suite of machine learning classifiers for sorting. The workflow concludes with model assessment employing k-fold cross-validation and statistical significance probing, followed by graphical representation, thereby delivering precise discrimination and informative insights into rice kernel varieties.

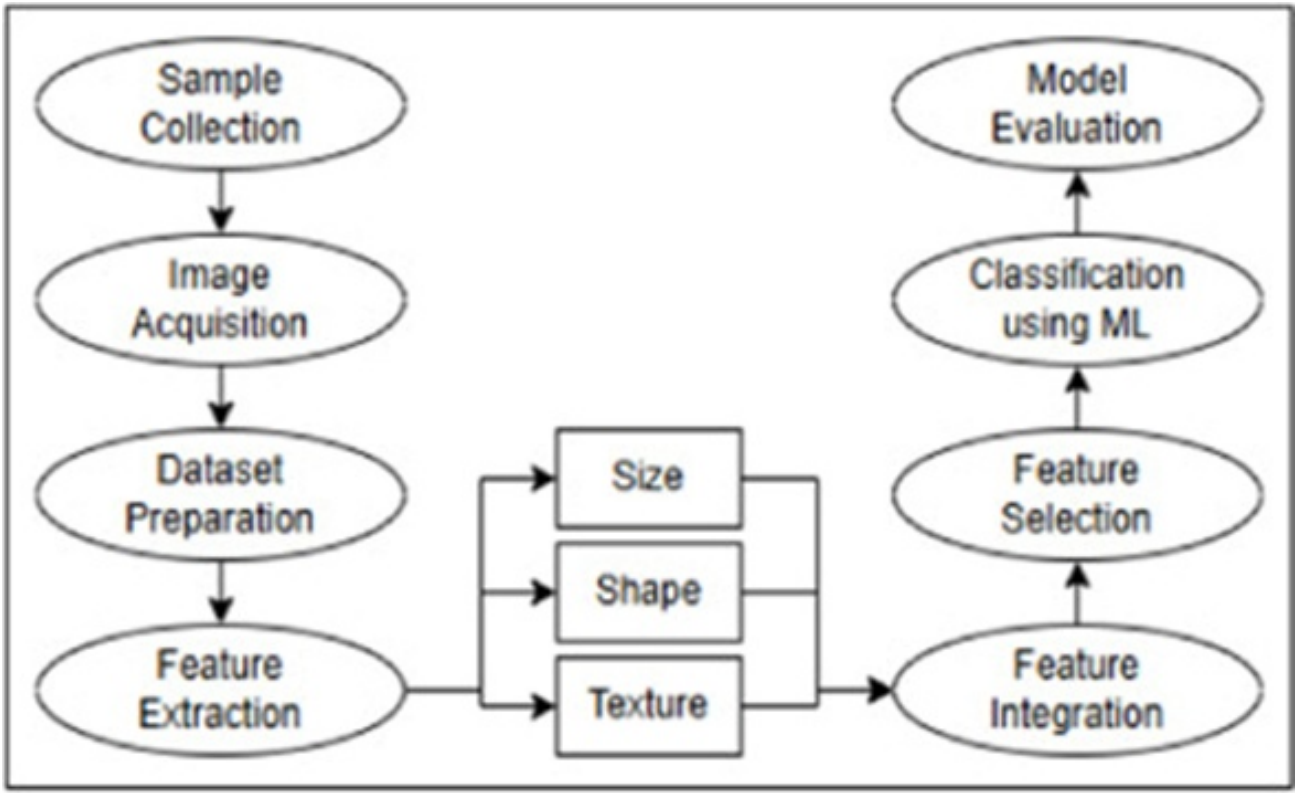

Figure 1. A block diagram depicting the overall methodology.

Sample collection

The grain specimens employed in this research were procured from the Zonal Agricultural and Horticultural Research Station, Brahmavar. The collection encompassed 300 rice kernels, split into 100 grains apiece for the three discrete

© 2022 The Author(s). | Published by Agro Intelligence Research Press

varieties: Uma, Jyothi, and KCP-1. Each accession was selected for its widespread cultivation and agronomic significance within the region, providing a varied substrate for comparative scrutiny of kernel traits.

Image acquisition

Photographing the rice kernels was conducted under controlled conditions to maintain uniformity across the entire sample pool. A 45-megapixel sensor was pressed into service, and the captures were taken beneath D-65 illuminant, which furnishes stable lighting. A backing of black velvet fabric was deployed to obviate shadows and reflective glints. The non-reflective nature of velvet proves indispensable, as it suppresses light scattering and ensures even illumination across every frame. This configuration, shown in **Figure 2**, facilitates the capture of high-quality imagery, which is essential for granular image dissection.

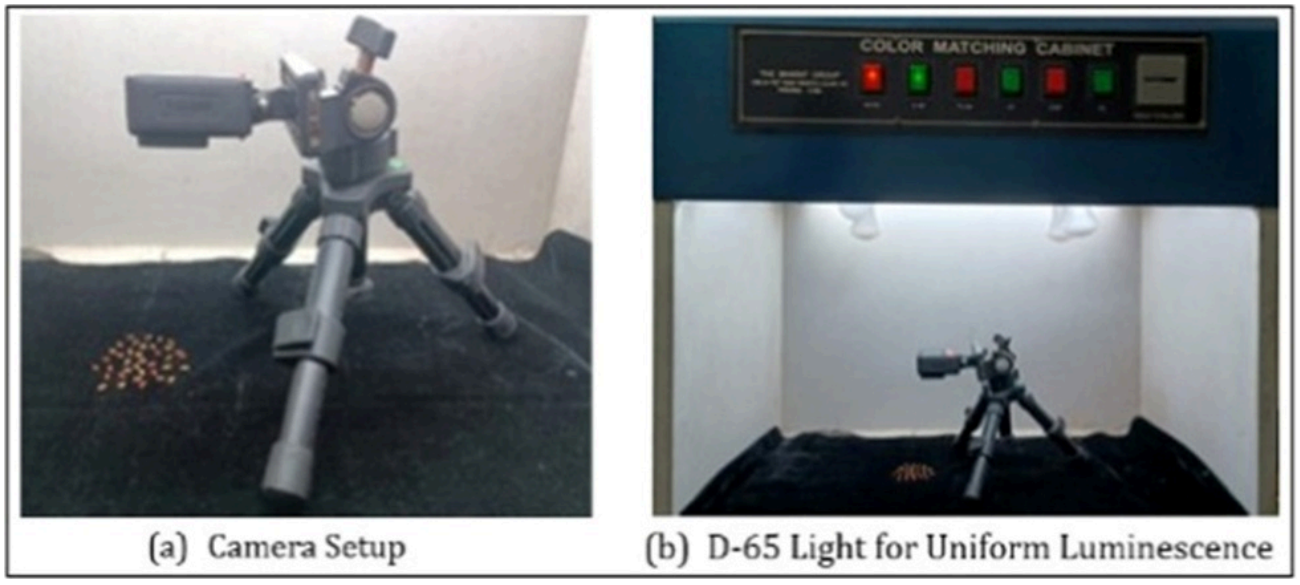

Figure 2. Image acquisition setup.

Dataset preparation

The dataset grooming pipeline was derived from the raw photograph shown in **Figure 3a**, organized into sequential stages to condition the 300 specimens—100 grains each of Uma, Jyothi, and KCP-1—for exhaustive descriptor extraction. The dataset was subsequently split into training and test sets, with 75% allocated to training and 25% to testing.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

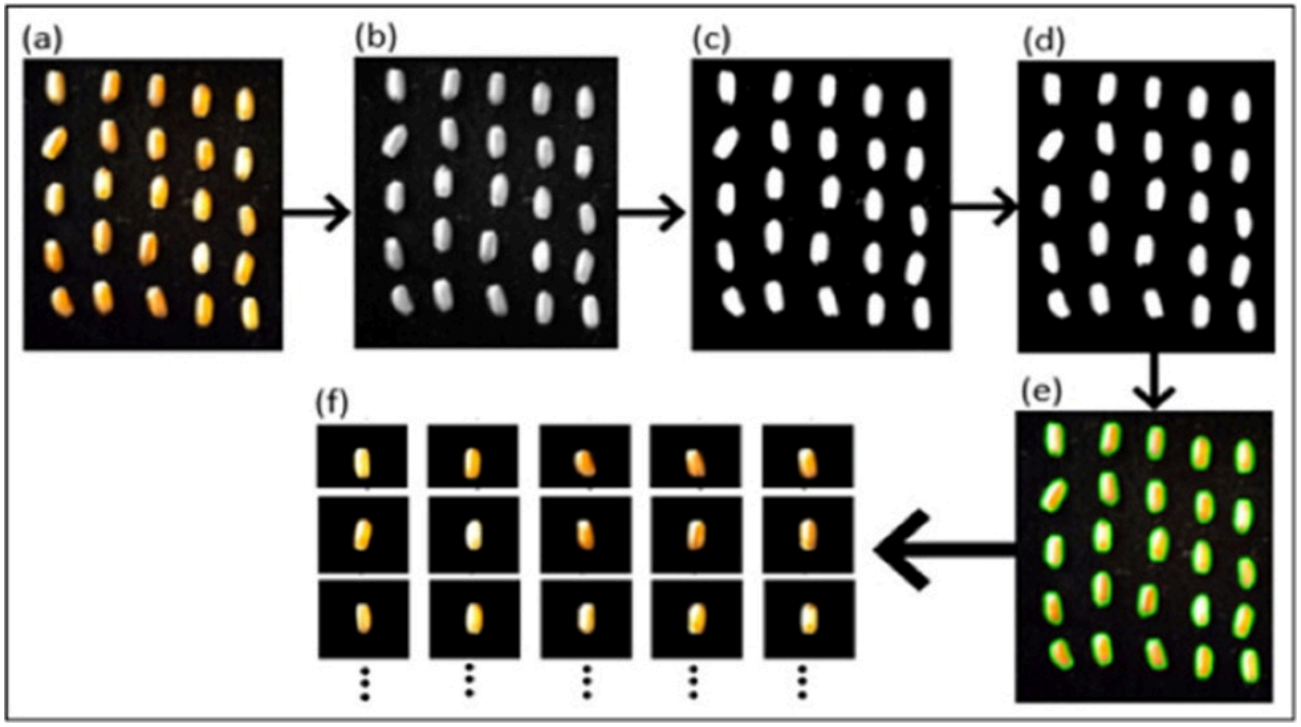

Figure 3. Steps involved in dataset preparation: (a) original image, (b) grayscale image, (c) thresholded image, (d) opened image, (e) identified contours, and (f) cropped rice grains.

Pre-processing

Every photograph was initially converted from the RGB colour space to grayscale to streamline the analytical workflow by reducing computational demands while preserving the requisite textural granularity. This mapping is formally captured by **Eq. 1**, in which the grayscale magnitude is a weighted sum of the colour band intensities.

$$\mathrm{Gray} = 0.299\mathrm{R} + 0.587\mathrm{G} + 0.114\mathrm{B} \quad (1)$$

Within this formulation, R, G, and B correspond to the magnitudes of the red, green, and blue channels, respectively. This conversion step is exemplified in **Figure 3b**. Thereafter, Otsu's thresholding (2) was applied to binarize the images. The procedure settles on an ideal cutoff that shrinks the within-category spread of the dark and light pixel populations, cleanly dissociating the kernels from the surrounding field, as evidenced in **Figure 3c**.

$$\sigma_{\mathrm{w}}^2(\mathrm{t}) = \omega_0(\mathrm{t})\,\sigma_0^2(\mathrm{t}) + \omega_1(\mathrm{t})\,\sigma_1^2(\mathrm{t}) \quad (2)$$

In that expression, ω0(t) and ω1(t) signify the probabilities of the two partitions carved out by the threshold t, while $\sigma^2_0(t)$ and $\sigma^2_1(t)$ denote the variances intrinsic to those partitions.

Noise removal

To ensure that the kernels stand unambiguously apart and are free of unwanted artefacts, morphological transformations such as opening were applied. This pivotal operation, steered by (3), scrubs the image by discarding minute-scale noise

© 2022 The Author(s). | Published by Agro Intelligence Research Press

and uncoupling grains that happen to be touching. These corrections are fundamental to the trustworthy detection of contours downstream. The progression from the partitioned depiction in **Figure 3c** to the cleaned depiction in **Figure 3d** vividly demonstrates the effectiveness of this noise-purging stage.

$$Opening\left(Img\right) = Dilation\left(Erosion\left(Img, S\right), S\right) \quad (3)$$

In the above, Img denotes the partitioned image, a binary canvas on which morphological operations are applied to purify and segment pictorial structures. S stands for the structuring template, a 4 × 4 array of ones, which steers how the erosion and dilation processes reshape the canvas Img.

Contour detection

Boundary traces of single kernels were extracted from the cleansed renditions. Charting the outline of each grain is indispensable for capturing both dimensional and form-related descriptors, as highlighted in **Figure 3e**.

Cropping and saving grains

Once contours had been pinpointed, the enclosing rectangles for each trace were calculated through (4), paving the way for the extraction of individual kernels. Every grain was then quarantined and archived as a standalone image. To guarantee that each excised kernel is placed dead-centre within a standardized frame, horizontal and vertical offset computations were performed as described in (5) and (6). This procedure underpins independent descriptor derivation and sharpens the accuracy of the collated data, as rendered in **Figure 3f**.

$$\left(x, y, w, h\right) = boundingRect\left(C_i\right) \quad (4)$$

Eq. 4 yields the coordinates x and y of the top-left vertex, alongside the breadth w and height h of the minimal bounding rectangle enclosing the i-th contour Ci.

$$\mathrm{X_{offset}} = \frac{(256 - w)}{2} \quad (5)$$

$$\mathrm{Y_{offset}} = \frac{(256 - h)}{2} \quad (6)$$

Eqs. 5 and 6 compute the translational offsets required to centre the trimmed kernel pictures inside a new 256×256 pixel frame, ensuring uniformity and repeatability throughout image evaluation.

Feature extraction

The descriptor extraction chain used in this work extracts a wide range of attributes from rice grains, with a focus on size, shape, and textural features. The sequence begins with the raw photograph, transposed to grayscale, then thresholded for partitioning, as shown in **Figures 4a-4c**, respectively. Textural descriptors were sourced from grayscale renditions because they rely on intensity modulations that remain intact in monochrome imagery. In contrast, size and shape

© 2022 The Author(s). | Published by Agro Intelligence Research Press

represent structurally grounded properties contingent upon the kernel silhouette and were consequently drawn from the partitioned depictions.

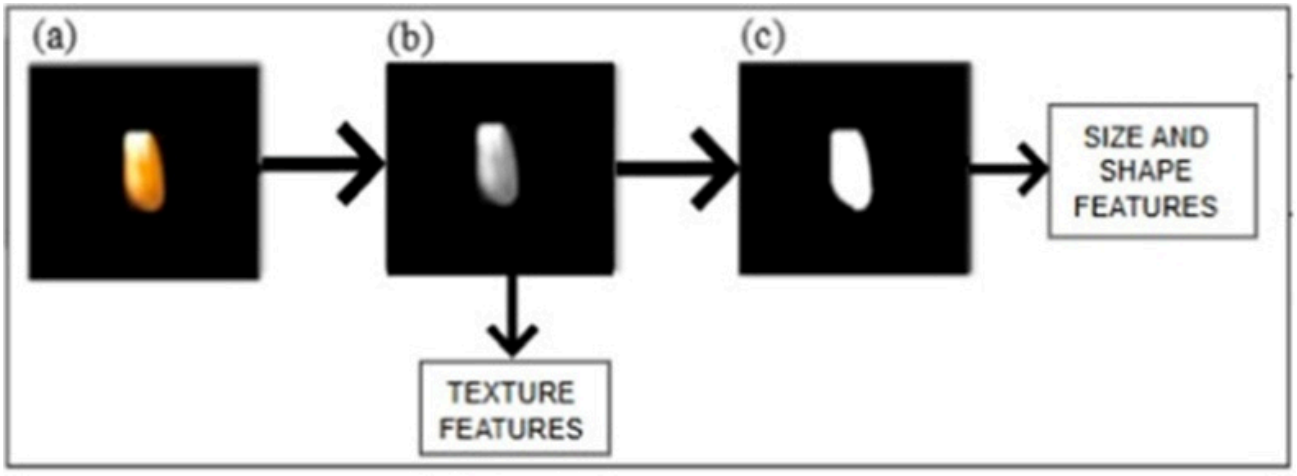

Figure 4. Steps Involved in Dataset Preparation: (a) original image, (b) grayscale image, (c) segmented image, and (d) labelled image.

Across this study, separate descriptor divisions are harnessed to assess the innate qualities of rice varieties. Features F1 through F10, tabulated in **Table 1**, are categorized as textural attributes, pivotal for appraising surface characteristics that distinguish rice varieties. Features F11 to F16 zero in on dimensional metrics, fundamental for mapping the physical boundaries of the grains. Features F17 to F25 investigate morphological properties, delivering fine-grained perception into the structural subtleties of rice kernels.

Table 1. List of features extracted.

Feature code	Feature title	Reworded description
F1	Mean	A statistical indicator representing variations in brightness and contrast across the grain image.
F2	Standard Deviation	A statistical parameter reflecting the distribution and spread of pixel intensity values.
F3	Variance	A statistical metric used to determine the extent of dispersion among pixel intensities.
F4	Skewness	Represents the degree of asymmetry in the grayscale intensity distribution.
F5	Kurtosis	Represents the peakedness and tail-distribution characteristics of grayscale intensities.
F6 [0–9]	Local Binary Pattern (LBP) Histogram	Computed using the Uniform LBP approach with the following settings: Radius = 1, Number of Neighbours = 8, and Method = ‘uniform’. This descriptor characterizes local texture by measuring the occurrence frequency of uniform binary patterns within neighbouring pixels.
F7	Contrast	Evaluates local intensity variations in the grayscale image based on gradient magnitude.
F8	Directionality	Describes the dominant orientation of image gradients.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

F9	Edge Density	Refers to the ratio of detected edge pixels after edge filtering, providing information about the smoothness or roughness of the grain boundary.
F10	Texture Energy	Calculated as the sum of squared values in a smoothed image, reflecting texture consistency and uniformity.
F11	Length	The grain dimension is measured along its maximum longitudinal axis.
F12	Width	The transverse measurement of the grain is taken perpendicular to its length.
F13	Aspect Ratio	The proportional relationship between grain length and width indicates grain elongation.
F14	Area	The complete number of pixels occupied by the grain region.
F15	Perimeter	The total boundary length surrounding the grain object.
F16	Lack of Shape	Measures the irregularity of the grain form relative to an ideal rectangle, aiding in differentiating rice varieties through subtle structural variations. It is determined using: F16 = (F11 × F12) / F14
F17 [[1], [2], [3], [4], [5], [6], [7]]	Hu Moments	Shape descriptors are invariant to geometric transformations such as translation, scaling, and rotation, enabling robust shape characterization.
F18	Eccentricity	Quantifies the deviation of the grain shape from a perfect circular form.
F19	Solidity	The proportion between the grain area and the area of its convex hull represents grain compactness.
F20	Extent	The ratio between the pixels contained within the grain region and those inside the bounding box.
F21	Orientation	Specifies the angle of grain alignment relative to the horizontal axis.
F22	Major Axis Length	The maximum possible line length passing through the grain structure.
F23	Minor Axis Length	The shortest line dimension perpendicular to the major axis passing through the grain.
F24	Compactness	A shape descriptor indicating how closely the grain resembles a circle with an equivalent perimeter.
F25	Convexity	Describes the degree of curvature or outward bulging of the grain surface, which is beneficial for distinguishing grains with irregular boundaries among rice varieties. It is calculated as: F25 = F15 / Convex Perimeter, where “Convex Perimeter” refers to the perimeter of the convex hull surrounding the grain, representing the smallest convex boundary enclosing all outer points while excluding indentations.

Feature integration

Feature integration constitutes a keystone of the methodological framework, entailing the fusion of disparate descriptor families—size, shape, and texture—into a single amalgamated dataset. This merging enables interrogation of every plausible permutation of these descriptor clusters, yielding seven distinct scenarios that span all conceivable single- and combined-feature arrangements. Such a tactic lays the foundation for a comparative inquiry to identify which assemblies most effectively enhance classification fidelity for rice grain varieties. By methodically dissecting these permutations, the most revealing descriptors can be pinpointed, thereby refining ML models for precise and efficient grain type separation.

Feature selection

© 2022 The Author(s). | Published by Agro Intelligence Research Press

Feature selection enhances the performance of machine learning models by isolating and safeguarding the most consequential features. Recursive Feature Elimination (RFE) was adopted for SVM, RF, and LR because these models produce feature importance scores or coefficient magnitudes, enabling attributes to be ranked and the least impactful ones to be iteratively removed. In the case of KNN and GNB, which are devoid of native feature ranking faculties, Backward Feature Elimination (BFE) was instituted by discarding attributes singly while observing performance fluctuations. These tactics reduced model complexity, reduced computational costs, and improved classification accuracy by jettisoning superfluous or irrelevant descriptors. The mechanics of RFE and BFE are respectively sketched in **Figures 5 and 6**. Within these schematics, S captures the active descriptor pool, fk **(Figure 5)** and fi **(Figure 6)** denote single attributes together with their ranking position. N signals the desired count of attributes to be retained upon completion of the culling cycle.

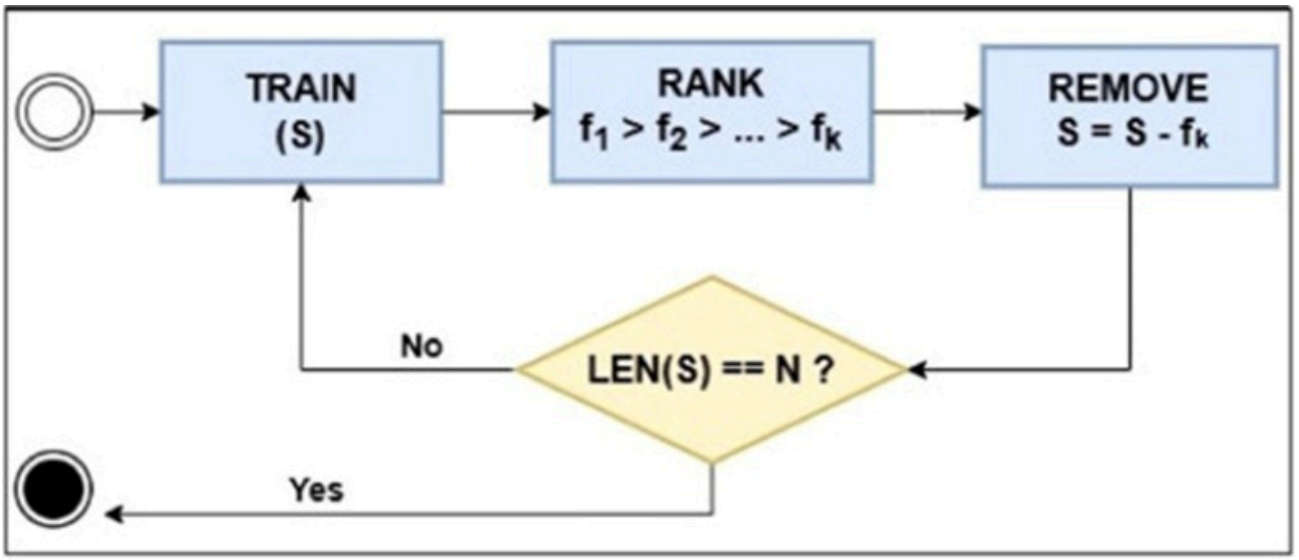

Figure 5. Recursive Feature Elimination Process.

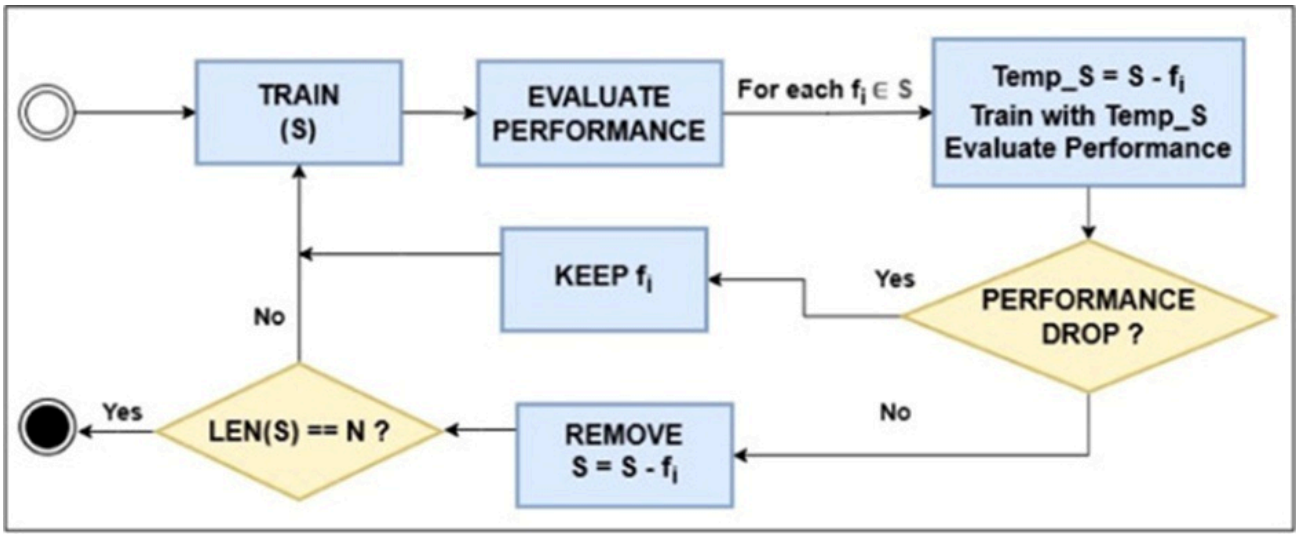

Figure 6. Backward Feature Elimination Process.

Classification using ML models

© 2022 The Author(s). | Published by Agro Intelligence Research Press

Throughout this study, an assortment of ML models spanning Support Vector Machines (SVM), Random Forest (RF), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), K Nearest Neighbours (KNN), and Ensemble Learning (EL) methodologies were enlisted to sort rice grains based on the extracted descriptors. Every model was picked for its distinctive aptitudes in addressing varied dimensions of the taxonomic challenge.

Support vector machine

SVMs excel at finding an optimal separating hyperplane that maximizes the buffer zone between categories. This quality is especially useful for datasets whose classes are linearly separable, making SVMs particularly well-suited for crisp, well-delineated classification assignments. The SVM mandate of margin maximization is codified as max_(w,b)[f0 〖 =1/(|(|w|)|)〗 , where w denotes the weight vector that governs the hyperplane's spatial attitude, and b represents the offset term enabling the hyperplane to translate, thereby perfecting the separation interval between classes.

$$P(y|x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i|y)}{P(x_1, \ldots, x_n)} \quad (7)$$

Random forest

RF is harnessed for its built-in resistance to overfitting, a recurrent vulnerability in high-capacity models. By combining an ensemble of decision trees, RF not only delivers robust predictive performance but also provides informative feature relevance metrics. These indicators are instrumental in surfacing the most telling descriptors within disorderly or entangled datasets, thereby buttressing model dependability.

$$P(y = 1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\ldots+\beta_n x_n)}} \quad (8)$$

Gaussian Naïve Bayes

GNB is favoured on account of its parsimony and potency, especially under circumstances where the premise of predictor independence holds valid. It is computationally parsimonious and performs respectably in multiclass forecasting contexts, making it a suitable candidate for tackling large datasets. The Gaussian Naïve Bayes model estimates the conditional likelihood of label y given predictors x_1, …, x_n, under the assumption of mutual independence among the predictors. This probabilistic scaffold, which underpins the classification verdict, is outlined in (7).

Logistic regression

LR is well-equipped to address multiclass discrimination problems through a one-versus-rest formulation. This design makes it notably powerful in situations involving multiple target categories, as encountered when distinguishing among disparate rice grain types. Its built-in ability to generate class-conditional probability estimates strengthens its interpretability, an asset that lends itself to fine-grained, class-by-class examination. The logistic function, encoded in (8), estimates the chance that a given input vector x is associated with the positive outcome y = 1.

K-Nearest neighbours

KNN earns its place for its conceptual simplicity and practical efficacy, showing particular value where the separating boundary between categories remains ill-defined. The algorithm rests on a principle of locality, in which the label assigned to a query instance is determined by the majority vote of its closest neighbours. This property allows KNN to navigate datasets characterized by convoluted, non-linear class boundaries with considerable skill, grounding each forecast in the contextual evidence furnished by the most proximate observations.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

Ensemble learning

Ensemble Learning, instantiated here through a stacking architecture, combines several component models to leverage their collective strengths while tempering their respective frailties. This blueprint, laid out in **Figure 7**, enhances both the predictive accuracy and the robustness of the classification system, making it supremely reliable for complex decision-making problems. In the present investigation, the stacking construct incorporated LR, RF, SVM, and GNB as base-tier learners, with GNB serving as the final-stage aggregator that coalesces the individual predictions into a single coherent verdict. GNB was selected for this meta-learner function after experimenting with a variety of classifiers, as it repeatedly produced the strongest validation scores—quite conceivably reflecting its minimalism, computational speed, and ability to handle the heterogeneous meta-variables generated by the constituent models.

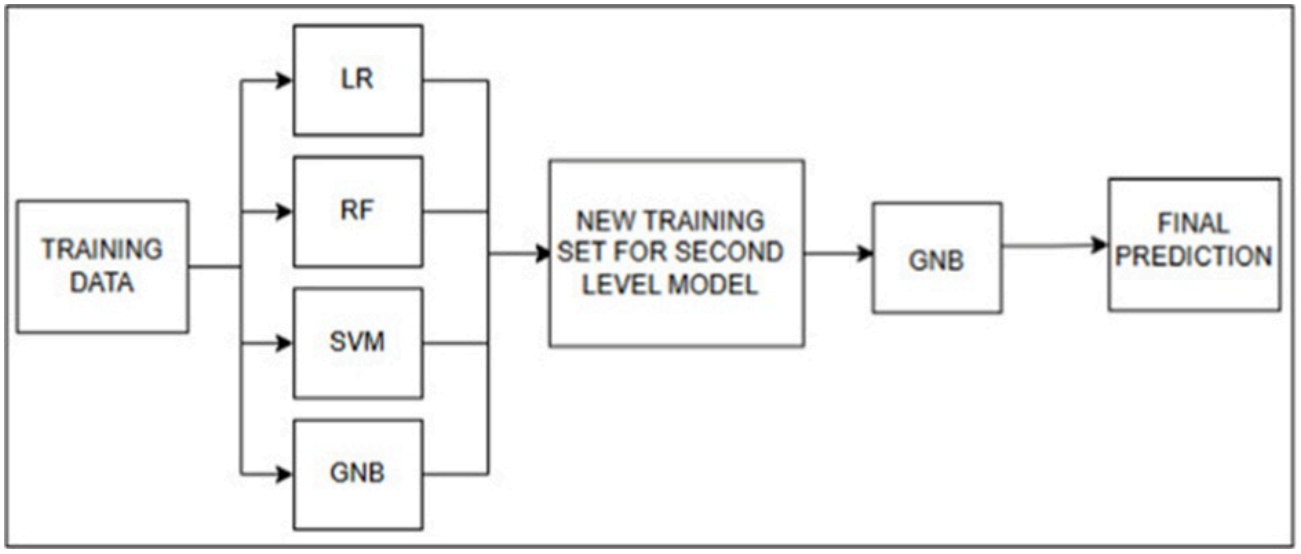

Figure 7. Workflow of stacking ensemble model.

Model evaluation

This section describes the techniques used to measure model effectiveness. A collection of evaluative indicators was used, including accuracy, precision, recall, f1-score, confusion matrices, receiver operating characteristic (ROC) graphs, and area under the curve (AUC), to provide a well-rounded picture of classification performance.

Performance metrics

The chief benchmarks used to evaluate the models include accuracy, precision, recall, and the F1-score, as formalized in **Eqs. 9-12**. Accuracy reflects the global proportion of correct assignments spanning all categories. In contrast, precision and recall offer per-class visibility into model conduct, an aspect of prime importance in contexts where the cost imbalance between false positives and false negatives cannot be ignored. The F1-score has particular utility as the harmonic mean of precision and recall, combining the two into a single metric that captures their balance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

© 2022 The Author(s). | Published by Agro Intelligence Research Press

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

Confusion matrix

A confusion matrix, shown in **Figure 8**, is an indispensable diagnostic tool for understanding the behaviour of a classification engine. It sorts the predicted labels into four compartments: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This grid helps visualize prediction fidelity class by class, highlighting not only the hits (TP and TN) but also the blunders—FP, where the model wrongly flags the affirmative class, and FN, where it overlooks it entirely. Such granular inspection enables researchers to discern recurring misclassification motifs, offering a roadmap for sharpening their categorization tactics.

		Prediction Class				
		C1	C2	C3		Cn
Actual class	C1	T1	F12	F13	...	F1n
	C2	F21	T2	F23	...	F2n
	C3	F31	F32	T3	...	F3n
	...	...	...	...	...	...
	Cn	Fn1	Fn2	Fn3	...	Tn

Note: C: Class, T:True, and F: False

Figure 8. General confusion matrix for multiclass classification.

ROC and AUC metrics

The receiver operating characteristic (ROC) curve and its area under the curve (AUC) are used to assess model performance across a continuum of decision cutoffs. The ROC plot charts the true positive rate as a function of the false positive rate, illuminating the trade-off between correctly identifying hits and absorbing false alarms. The AUC condenses this entire profile into a single numeric summary; the closer the value approaches unity, the better the model's ability to separate the competing classes.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

Experimental Findings and Results

A thorough-going analysis was performed across all seven feature permutations using six ML architectures, scrutinizing their output via accuracy, precision, recall, F1-score, confusion matrices, and feature selection footprints. In the interest of brevity and readability, while **Tables 2 and 3** present the accuracy tallies juxtaposed across the full complement of feature groupings, the fine-grained performance readouts are spotlighted exclusively for the Size + Shape + Texture configuration, which captured the highest mean accuracy of 96.44 %, as visualized in **Figure 9**, across the suite of classifiers.

Table 2. Accuracy comparison across individual feature sets.

Empty cell	Texture	Shape	Size
SVM	80	92	90.67
RF	85.34	92	90.67
GNB	82.67	86.67	94.67
LR	90.67	92	93.34
KNN	94.67	94.67	93.34
EL	90.67	92	93.34

Table 3. Accuracy comparison across combined feature sets.

Empty cell	Size, shape, and texture	Shape and texture	Size and texture	Size and shape
SVM	96	96	90.67	97.34
RF	93.34	93.34	93.34	96
GNB	96	96	90.67	96
LR	97.34	98.67	96	92
KNN	98.67	95	98.67	96
EL	97.34	97.34	96	93.34

© 2022 The Author(s). | Published by Agro Intelligence Research Press

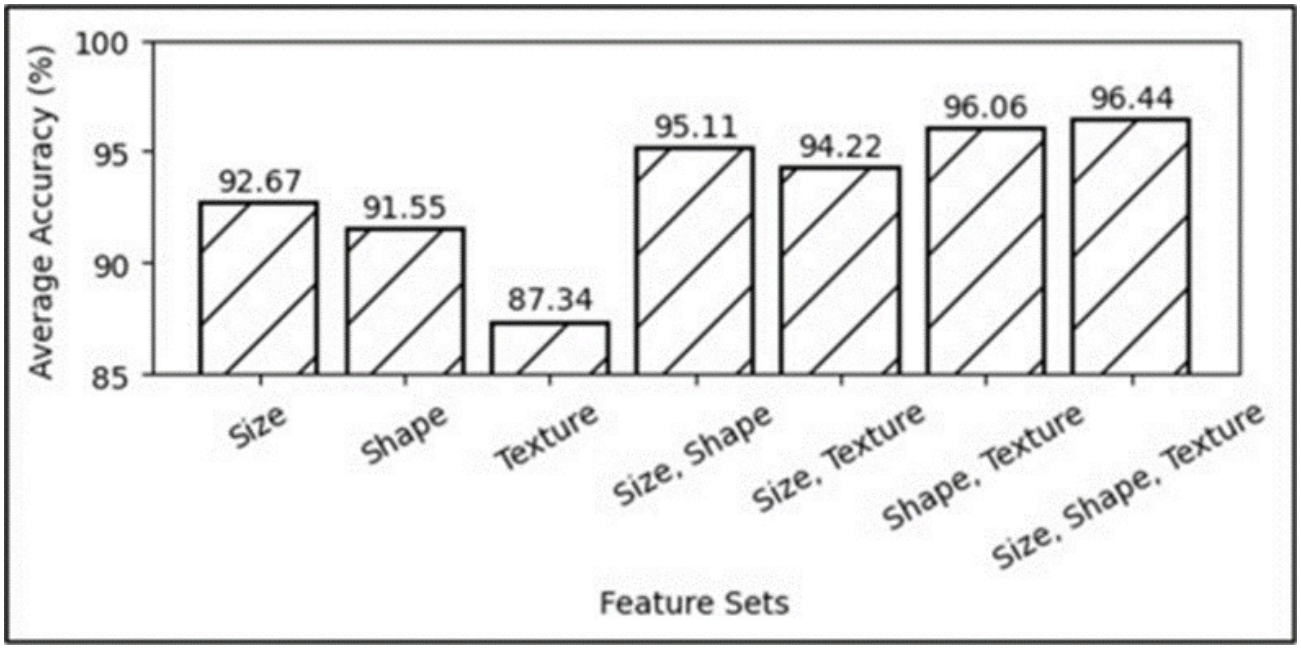

Figure 9. Average accuracy across models for various feature sets.

K-fold cross-validation: evaluating model generalization

To verify the extrapolative reach of the models and screen for signs of overfitting, k-fold cross-validation was implemented—a resilient scheme that partitions the dataset into k equally sized segments, or folds. At each round, k-1 folds constitute the training set, while the remaining fold serves as the validation set; the cycle repeats until every fold has served once as the evaluation partition. This regimen yields a panoramic assessment of model behaviour across disparate data slivers, conferring metrics that are considerably more trustworthy than those drawn from a one-off train–test division. The study at hand used k-fold cross-validation with k = 5, allowing each model to be evaluated on five distinct validation subsets. As shown in **Table 4**, the findings reveal tight agreement between cross-validation (CV) averages and test-set accuracies across all models, with discrepancies consistently within a single standard deviation. This congruence, paired with relatively contained standard deviations (most pronounced for LR and RF), indicates that the models maintain steady performance across diverse data partitions and exhibit no substantial overfitting. The front-running model, KNN, achieved a test accuracy of 98.67% compared with a cross-validation mean of 96.80%, suggesting robust generalizability despite its high accuracy.

Table 4. K-fold cross-validation results for classification models.

Model	Standard deviation	Mean CV accuracy (%)
SVM	3.44	95.56
RF	1.02	92.50
GNB	3.10	92.60
LR	1.09	96.44
KNN	3.00	96.80
EL	1.41	95.56

Performance analysis of individual feature sets

© 2022 The Author(s). | Published by Agro Intelligence Research Press

When examining the standalone descriptor groupings, as itemized in **Table 2**, marked differences in model behaviour emerged across the different machine learning techniques. The accuracy figures logged for the solitary feature families—Size, Shape, and Texture—attest that each contributes a distinctive component to the sorting pipeline. The Size attribute bundle consistently yielded higher accuracy across the full model spectrum, with a mean accuracy of 92.67%, indicating that dimensional cues carry appreciable discriminatory weight for the classifiers employed. Shape descriptors likewise put up a strong showing, attaining a mean accuracy of 91.55%, with the KNN algorithm achieving the highest value. This bears out the potency of morphological indices in teasing apart classes within a spatial landscape. By comparison, textural measures exerted the least influence on model accuracy, averaging 87.34%, suggesting that surface texture, taken alone, may be less distinguishing for the algorithms under evaluation than size and shape. This portion of the analysis highlights the importance of careful feature curation in machine learning and gestures toward the potential dividends of descriptor fusion as a route to sturdier, more accurate models.

Efficacy of combined feature sets

The investigation focuses on the performance gains unlocked by blending composite descriptor ensembles into machine learning workflows. The evidence arrayed in **Table 3** supplies a side-by-side comparative breakdown across a range of classifiers—SVM, RF, GNB, LR, KNN, and EL—exploiting couplings such as Size + Shape, Size + Texture, Shape + Texture, and Size + Shape + Texture. This head-to-head exposition throws into sharp relief the considerable leverage that feature amalgamation wields over model fidelity. The results make plain that fusing descriptor families routinely outstrips the performance attainable with single-source feature sets. The Size + Shape + Texture triad emerges as the pre-eminent combination, securing the top-tier mean accuracy of 96.44%. Tellingly, the KNN classifier achieved 98.67% accuracy with this descriptor configuration, highlighting the immense promise of leveraging comprehensive feature portfolios in complex classification settings. Alongside this, the Shape + Texture duo also returns encouraging figures (96.06% mean accuracy across models), lying within close striking distance of the leading mean accuracy and underscoring the constructive interplay between these two descriptor classes. This observation suggests that even selective fusions of feature sets can encapsulate pivotal discriminatory content, substantially magnifying classifier performance.

Detailed analysis of optimal feature set

The feature curation phase uncovered model-specific learnings concerning the most favourable descriptor subsets. As shown in **Table 5**, certain classifiers—GNB and EL—performed best when supplied with the unabridged feature collection, whereas others achieved superior efficiency with bespoke, pared-down descriptor groupings. As **Table 5** demonstrates, SVM and RF achieved their highest efficacy with compressed feature sets, while KNN demonstrated exceptional classification performance with a slender selection of judiciously chosen descriptors. Such heterogeneity in feature predilections across models underscores the need for classifier-specific descriptor optimization rather than a monolithic, blanket methodology.

Table 5. Model performance and selected features for optimal feature set.

Selected attributes	Classification model	Evaluation metric	Jyothi	KCP-1	Uma
F13, F17 [3], F17 [4], F21, F22, F23, F1, F2, F5, F7, F10	SVM	Precision	100	100	89
		Recall	92	96	100
		F1-Score	96	98	94
F11, F12, F13, F14, F17 [1], F17 [2], F17 [3], F17 [4], F18, F21, F11, F22, F23, F11, F1, F2, F3, F4, F5, F6 [3], F6 [5], F6 [8], F7, F9, F10	RF	Precision	95	89	96

© 2022 The Author(s). | Published by Agro Intelligence Research Press

		Recall	84	96	100
		F1-Score	89	92	98
All Features	GNB	Precision	96	92	100
		Recall	93	96	100
		F1-Measure	94	94	100
F11, F18, F2, F6 [6], F17 [1], F17 [2], F17 [4], F17 [5], F17 [7], F19, F20, F21, F1, F2, F6 [0], F9	LR	Precision	93	100	100
		Recall	100	92	100
		F1-Score	96	96	100
F11, F13, F14, F16, F6, F7, F9	KNN	Precision	96	100	100
		Recall	100	96	100
		F1-Score	98	98	100
All Features	EL	Precision	93	100	100
		Recall	100	92	100
		F1-Score	96	96	100

The Size + Shape + Texture ensemble notched the highest mean accuracy of 96.44% across the board **(Table 3)**, plausibly owing to its capacity to marshal an expansive repertoire of distinguishing traits—dimensional attributes for physical scale, structural metrics for morphological subtleties, and surface indices for textural patterning—thereby enabling dependable demarcation among the Uma, KCP-1, and Jyothi cultivars. By contrast, the Size + Shape dyad shone particularly for SVM (97.34% accuracy, **Table 3**), conceivably because of SVM's facility in modelling both rectilinear and curvilinear associations within a more economical feature arena, where size and shape cues alone deliver adequate discriminatory leverage without the supplemental overhead of texture information, which could potentially inject spurious variability for this specific learner.

Probing the performance tallies in **Table 5** across the trio of rice accessions yields edifying insights into classification strength. As shown in **Table 5**, for the Uma cultivar, nearly all classifiers achieved remarkably elevated precision and recall, with several models reaching flawless ceilings (100%). For the KCP-1 cultivar, as shown in **Table 5**, SVM, LR, KNN, and EL all achieved perfect precision (100%), though recall ranged from 92% to 96%. Turning to the Jyothi cultivar, **Table 5** shows that precision remained stable across all models (93%- 100%), whereas recall showed the widest dispersion, with RF dipping to 84% and other contenders reaching 100%. The F1-scores enumerated in **Table 5**, which harmonize precision and recall, lend further weight to the soundness of the honed descriptor sets, with KNN delivering the most consistent F1-score profile (98%–100%) across all three cultivars. KNN's stellar track record with a trimmer feature palette **(Table 5)** may well stem from its foundation in proximity-based inference, wherein a concise nucleus of highly diagnostic descriptors curtails miscalculation in distance determination, thereby bolstering accuracy across cultivars. The durable, high-performance figures chronicled in **Table 5**, evident across a range of evaluative metrics and diverse rice accessions, attest to the transferability of the pinpointed optimal feature configurations for red rice taxonomy.

Hyper-parameter tuning

Hyperparameter refinement was performed to improve classification performance. In the case of SVM, the penalty magnitude, kernel family, and scale parameter were systematically tuned, ultimately selecting a polynomial kernel with

© 2022 The Author(s). | Published by Agro Intelligence Research Press

elevated regularization to carve out more incisive decision boundaries. RF calibration addressed the forest size, maximum allowable depth, minimum split threshold, and bootstrap participation, striking a poised equilibrium between expressive capacity and extrapolative reliability. GNB adjustments focused on the smoothing term to better handle low-dispersion conditions. LR was streamlined by tuning the regularization coefficient, penalty type, and numerical solver, while using a sparsity-promoting regularizer at a moderate level. KNN tuning addressed neighbourhood cardinality, distance formulation, and neighbour weighting, ensuring a well-regulated classification landscape. The premier hyperparameter settings for each model, together with a glossary of each term, are collated in **Table 6**.

Table 6. Hyper-parameter tuning.

Model	Optimized hyperparameters	Explanation of parameters
SVM	C = 100, kernel = poly, gamma = scale	Regularization coefficient, selected kernel function, and gamma value controlling the influence range of training samples
RF	n_estimators = 200, max_depth = 10, min_samples_split = 2, min_samples_leaf = 1, bootstrap = True	Total number of decision trees, maximum tree depth, minimum samples required for node splitting, minimum samples required at leaf nodes, and use of bootstrap resampling
GNB	var_smoothing = 1e-09	Smoothing factor applied to stabilize calculations for features with very small variance
LR	C = 1, penalty = l1, solver = liblinear	Regularization strength, type of regularization penalty applied, and optimization algorithm used for model training
KNN	n_neighbors = 5, weights = uniform, P = 2	Number of neighbouring samples considered, weighting approach for neighbours, and distance measurement metric (p = 2 corresponding to Euclidean distance)

Statistical significance analysis

The statistical reliability of performance gaps among models was assessed using the Friedman test, which indicated substantial overall variation across the classifier pool ($\chi^2 = 19.971$, $P = 0.001$). Pairwise contrasts conducted via paired t-tests showed that KNN, the front-runner, meaningfully outpaced SVM ($P = 0.003$), RF ($P = 0.008$), and GNB ($P < 0.001$). That said, no statistically significant differences were detected between KNN and the immediate runners-up, LR ($P = 0.695$) and EL ($P = 0.156$). Appraisal of effect magnitude via Cohen's d unmasked large effects when setting KNN side by side with RF ($d = 1.919$) and GNB ($d = 1.377$), a medium effect in relation to EL ($d = 0.529$), a small effect opposite SVM ($d = 0.384$), and a trifling effect when juxtaposed with LR ($d = 0.160$). Collectively, these observations verify that KNN's pre-eminence is statistically well-founded and not attributable to stochastic drift.

Analysis of confusion matrices

Close inspection of the confusion matrices shown in **Figures 10a-10f** reveals telling insights into classification dynamics across the three rice types. The diagonal cells capture correctly assigned instances for each model: the SVM correctly classifies the full complement of 25 Uma exemplars without error, and the KNN classifier successfully classifies 24 Jyothi specimens. A survey of the off-diagonal entries reveals that the KCP-1 and Jyothi cultivars are more susceptible to mutual mislabelling, as shown in **Figure 10c**, **Figure 10**, and **Figure 10f**, where several models exhibit confusion between these two classes. In marked contrast, the Uma cultivar consistently registers negligible misassignment across the entire model suite, with Figure 10a-10f documenting only trivial instances in which Uma is conflated with alternative varieties. This

© 2022 The Author(s). | Published by Agro Intelligence Research Press

pattern intimates that Uma harbours more idiosyncratic morphological hallmarks that the curated descriptors latch onto effectively, whereas KCP-1 and Jyothi share a closer kinship within the descriptor representation space.

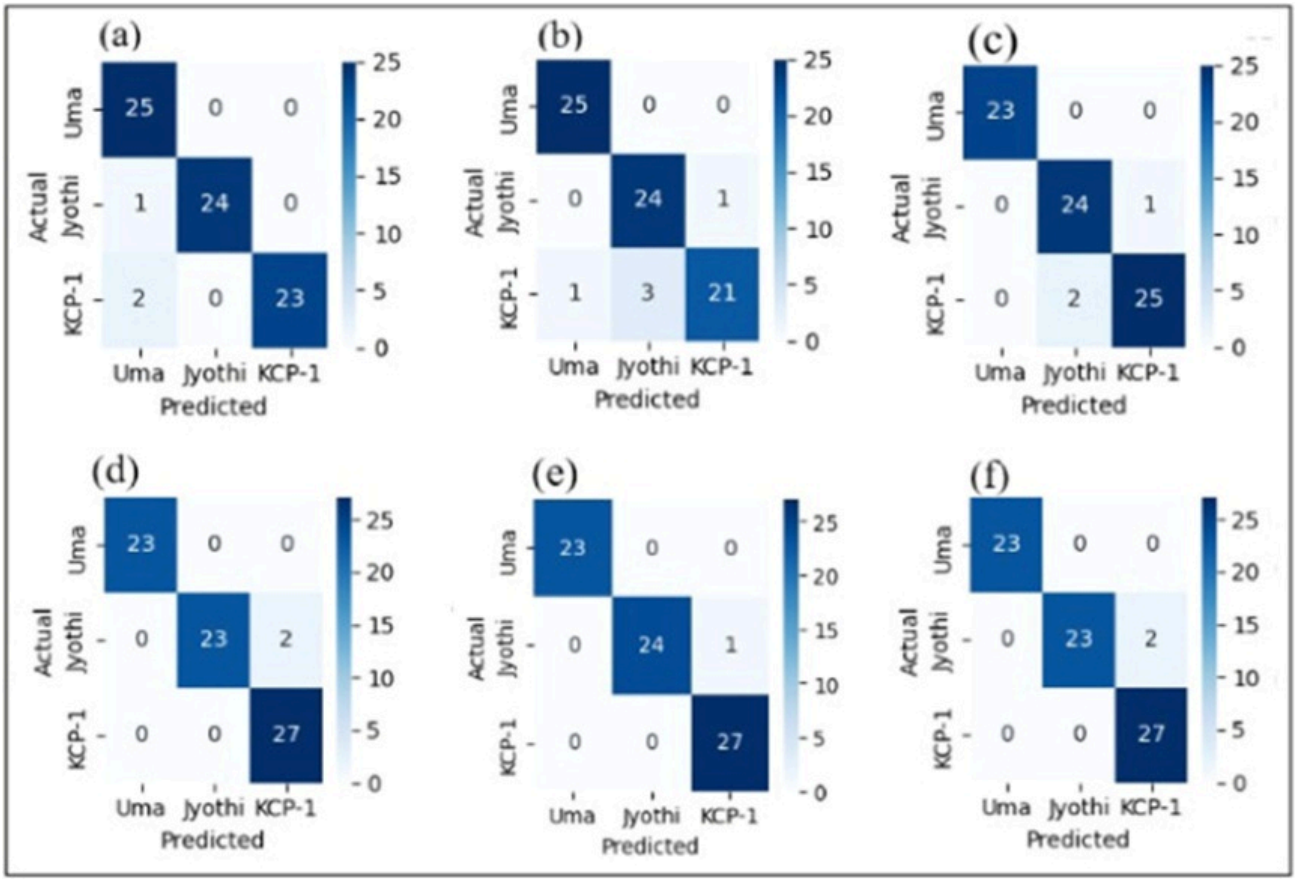

Figure 10. Confusion matrices: (a) SVM, (b) RF, (c) GNB, (d) LR, (e) KNN, and (f) EL.

Evaluation of ROC and AUC metrics

The ROC curves in **Figures 11a-11f** depict the classifiers' ability to distinguish among the categories across a continuum of decision thresholds by plotting the true positive rate against the false positive rate. The AUC statistic encapsulates this separation in a single figure; scores that converge toward 1.0 signal near-impeccable classification, whereas values lingering near 0.5 hint at performance barely surpassing random assignment. As observed throughout the panels, the models habitually achieve strikingly high AUC scores for all three rice accessions, with a profusion of curves either grazing or outright attaining the flawless classification benchmark of 1.0. The persistently sharp upward trajectory of the ROC curves toward the upper-left apex of each graph indicates that the classifiers maintain high sensitivity while suppressing false positives. This uniformly strong showing across the board endorses the efficacy of the chosen descriptor ensembles in distilling the signature traits of every rice accession. The near-spotless AUC scores, most conspicuously for Uma, bear out the dependable classification scaffold's capability to authoritatively separate the Uma, KCP-1, and Jyothi red rice cultivars.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

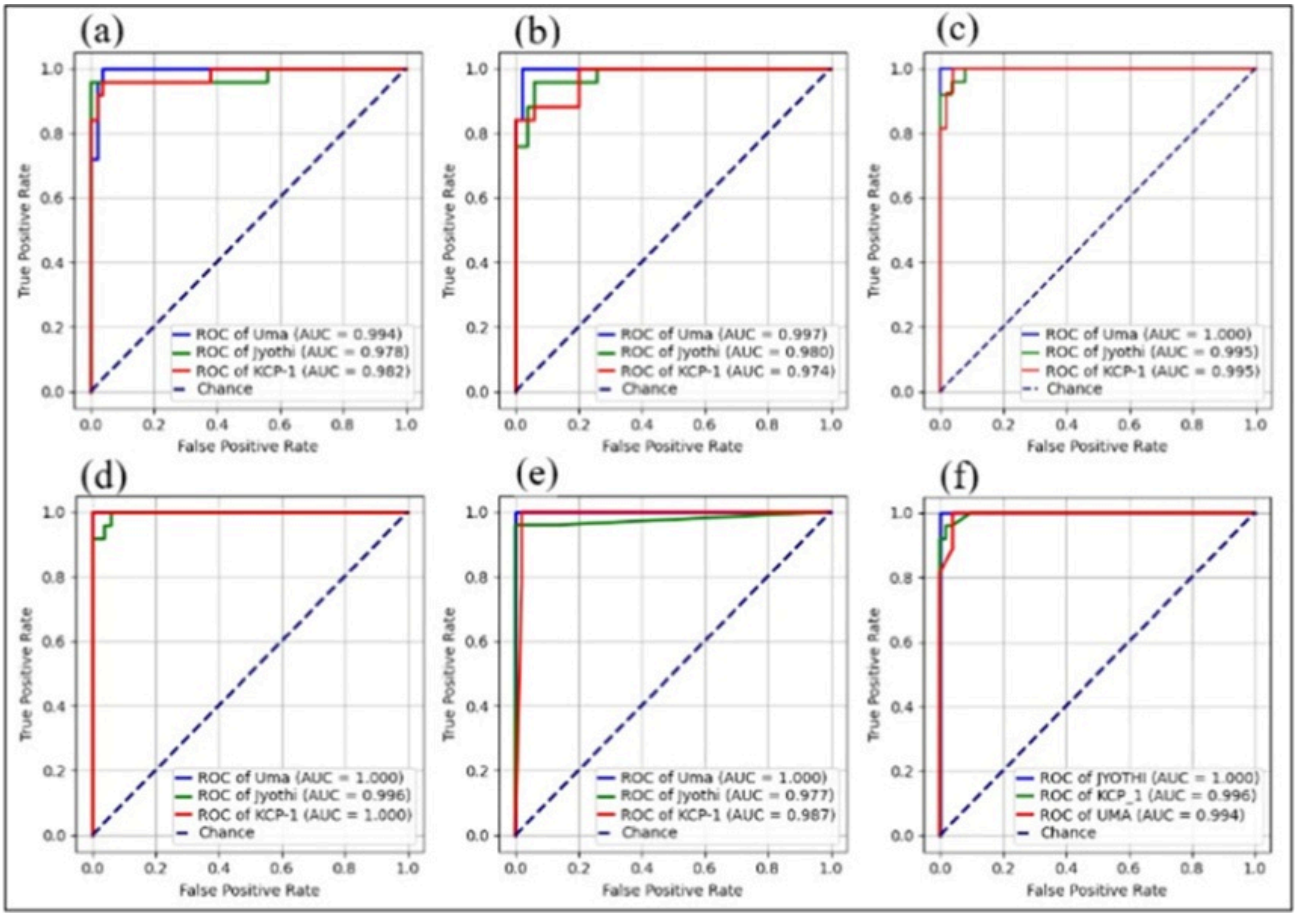

Figure 11. ROC curves and AUC metrics: (a) SVM, (b) RF, (c) GNB, (d) LR, (e) KNN, and (f) EL.

Conclusion

The present inquiry has productively carved out an original route toward red rice categorization, filling a pronounced lacuna in the research landscape where no earlier studies or benchmark collections devoted to red rice variety classification had been available. Through painstaking hyperparameter calibration across six disparate machine learning frameworks—SVM, RF, GNB, LR, KNN, and EL—combined with tactical descriptor winnowing via Recursive and Backward Feature Elimination protocols, the undertaking achieved exceptional classification results, with KNN reaching 98.67% accuracy. The fusion of size, shape, and texture descriptors proved to be the most commanding feature configuration, capturing a mean accuracy of 96.44% across the full model roster, albeit the shape combined with texture assembly trailed remarkably close, posting a mean accuracy of 96.06%. Exhaustive vetting, anchored in multiple indices, confirms that the study has reached a high level of classification excellence. The execution of k-fold cross-validation with k = 5 corroborated strong generalization across all models, without appreciable overfitting, thereby validating the potency of the optimized descriptor pairings. The multifaceted appraisal regime and the noteworthy performance yardsticks constitute a firm underpinning for automated red rice sorting infrastructures poised to upgrade agricultural quality governance procedures within the South Indian rice sector.

As for prospective investigations, several avenues of enrichment are put forward to further buttress and extend the purview of this contribution. Foremost, augmenting the dataset's cardinality is an imperative measure to fortify model resilience and its aptitude to generalize to more heterogeneous and previously unencountered specimens. Secondly, incorporating additional red rice landraces beyond Uma, Jyothi, and KCP-1 could help fashion a more inclusive and expandable taxonomic framework. Thirdly, once more copious datasets become available, deep learning paradigms

© 2022 The Author(s). | Published by Agro Intelligence Research Press

exemplified by Convolutional Neural Networks (CNNs) can be explored, as they may uncover representations of greater complexity and push classification accuracy beyond the thresholds achievable with conventional machine learning methodologies.

Acknowledgements

The authors are thankful to Dr. Sridevi Jakkaral, Assistant Professor with Specialization in Rice Breeding for Quality and Salinity, Department of Plant Science, Comprising Genetics and Plant Breeding, Plant Biotechnology, Crop Physiology, and Biochemistry, Zonal Agricultural & Horticultural Research Station, Brahmavar, Udupi District, Karnataka, India, for generously providing labelled rice grain varieties that were essential for our research work.

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Aug 2021 Revised: 19 Oct 2021 Accepted: 05 Nov 2021
Published online: 25 May 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. AICRIP: KCP-1. http://aicrip-intranet.in/Documents/AicripSite/Centers/14%20Brahmavar.pdf. Accessed 4 Feb 2025.

2. Bio basics: Uma Kerala red rice - parboiled. https://biobasics.org/products/uma-kerala-red-rice-parboiled?variant=37905433559199. Accessed 4 Feb 2025.

3. Husain AS. Impact of the Rice variety, Uma (Mo16) on farmers. Department of Agricultural Extension, College of Horticulture, Vellanikkara; 2015. M.S. thesis.

4. Arora B, Bhagat N, Saritha L, Arcot S. Rice grain classification using image processing & machine learning techniques. In: Proceedings of International Conference on Inventive Computation Technologies (ICICT); 2020. p. 205-8.

5. Nagoda N, Ranathunga L. Rice sample segmentation and classification using image processing and support vector machine. In: Proceedings of IEEE 13th International Conference on Industrial and Information Systems (ICIIS); 2018. p. 179-84.

6. Kiratiratanapruk K, Temniranrat P, Sinthupinyo W, Prempree P, Chaitavon K, Porntheeraphat S, et al. Development of paddy rice seed classification process using machine learning techniques for automatic grading machine. J Sensors. 2020;2020:1-15.

7. Aukkapinyo K, Sawangwong S, Pooyoi P, Kusakunniran W. Localization and classification of rice-grain images using region proposals-based convolutional neural network. Int J Autom Comput. 2020;17(2):233-46.

8. Krishna CV, Suchitra B, Helen S, Roobini MS. Quality analysis of rice grains using morphological techniques. In: Proceedings of International Conference on Communication, Computing and Internet of Things (IC3IoT); 2022. p. 1-5.

9. Ibrahim S, Zulkifli NA, Sabri N, Shari AA, Noordin MRM. Rice grain classification using multi-class support vector machine (SVM). IAES Int J Artif Intell. 2019;8(3):215-22.

10. Ruslan R, Khairunniza-Bejo S, Jahari M, Ibrahim MF. Weedy rice classification using image processing and a machine learning approach. Agriculture. 2022;12(5):645-58.

11. Singh KR, Chaudhury S. A cascade network for the classification of rice grain based on single rice kernel. Complex Intell Syst. 2020;6:321-34.

12. Cinar I, Koklu M. Identification of rice varieties using machine learning algorithms. J Agri Sci. 2022;28(2):9-17.

13. Farahnakian F, Sheikh J, Farahnakian F, Heikkonen J. A comparative study of state-of-the-art deep learning architectures for rice grain classification. J Agri Food Res. 2024;15:100890.

14. Nga TTK, Pham TV, Tam DM, Koo I, Mariano VY, Do-Hong T. Combining binary particle swarm optimization with support vector machine for enhancing rice varieties classification accuracy. IEEE Access. 2021;9:66062-78.

15. Kuo TY, Chung CL, Chen SY, Lin HA, Kuo YF. Identifying rice grains using image analysis and sparse-representation-based classification. Comput Electron Agric. 2016;127:716-25.

16. Carneiro LO, Coradi PC, Rodrigues DM, Lima RE, Teodoro LPR, Moraes RS, et al. Characterizing and predicting the quality of milled rice grains using machine learning models. AgriEngineering. 2023;5(3):1196-215.

17. Ahmed SB, Ali SF, Khan AZ. On the frontiers of rice grain analysis, classification and quality grading: a review. IEEE Access. 2021;9:160779-96.

18. Srimulyani W, Musdholifah A. Identification of rice variety using geometric features and neural network. Indonesian J Comput Cybern Syst. 2019;13(3):301-12.

19. Singh KR, Chaudhury S. Efficient technique for rice grain classification using back-propagation neural network and wavelet decomposition. IET Comput Vis. 2016;10(8):780-7.

20. Aznan A, Gonzalez Viejo C, Pang A, Fuentes S. Computer vision and machine learning analysis of commercial rice grains: a potential digital approach for consumer perception studies. Sensors. 2021;21(19):6354-70.

21. Hamzah AS, Mohamed A. Classification of white rice grain quality using ANN: a review. IAES Int J Artif Intell. 2020;9(4):600-8.

22. Sultana S, Faruque M, Islam MR. Rice grain quality parameters and determination tools: a review on the current developments and future prospects. Int J Food Prop. 2022;25(1):1063-77.

23. Ahad MT, Li Y, Song B, Bhuiyan T. Comparison of CNN-based deep learning architectures for rice diseases classification. Artif Intell Agric. 2023;9:22-35.

© 2022 The Author(s). | Published by Agro Intelligence Research Press

24. Tran-Thi-Kim N, Pham-Viet T, Koo I, Mariano V, Do-Hong T. Enhancing the classification accuracy of rice varieties by using convolutional neural networks. Int J Electr Electron Eng Telecom. 2023;12(2):150-60.

25. Patel B, Sharaff A. Rice variety classification and yield prediction using semantic segmentation of agro-morphological characteristics. Multimed Tools Appl. 2023;82(29):45567-84.

26. Kurade C, Meenu M, Kalra S, Miglani A, Neelapu BC, Yu Y, et al. An automated image processing module for quality evaluation of milled rice. Foods. 2023;12(6):1273-85.

27. Deepika C, Gnanamalar R, Thangaraj K, Revathy N, Karthikeyan A. Linear discriminant analysis of grain quality traits in rice (Oryza sativa L.) using the digital imaging technique. J Cereal Sci. 2023;109:103609.

28. Philip TM, Anita HB. Rice grain classification using Fourier transform and morphological features. Indian J Sci Technol. 2017;10(14):1-6.

© 2022 The Author(s). | Published by Agro Intelligence Research Press