

Machine Learning for Ripeness Classification of Nendra Bale Banana: A Classical Approach Using SVM, kNN, and Random Forest

MANASA T. N.¹, M. P. PUSHPALATHA²

^{1,2}*Department of Computer Science and Engineering, JSS Science and Technology University, Mysore, India*

Abstract- acceptance, and overall shelf life. Nendra Bale (Nendran), a dual-purpose, starch-rich Indian banana cultivar widely used for both table consumption and cooking, exhibits a ripening pattern that is markedly more texture-driven than colour-driven, distinguishing it from commercially dominant, colour-dominant cultivars. This study introduces an image-based framework for classifying the ripening stage of Nendra Bale bananas using classical machine learning. A first-hand dataset of 80 raw images across four ripening stages (Green, Mid-ripe, All-Yellow, Overripe) was collected under everyday mobile-camera conditions and augmented to 5,391 images to support more robust model training. Three classical classifiers (Support Vector Machine, k-Nearest Neighbours, Random Forest) were evaluated on Histogram of Oriented Gradients (HOG) features under both raw and augmented conditions. On the raw dataset, kNN gave the highest accuracy (76.47%), ahead of SVM and Random Forest (70.59% each); once the dataset was augmented, SVM became the strongest and most consistent model (81.60%), ahead of kNN (78.62%) and Random Forest (74.53%). A probabilistic stage-to-shelf-life mapping is proposed to convert predictions into an expected remaining life in days. The study offers a low-cost, non-destructive, deployable approach for quality assessment and post-harvest decision support for a texture-dominant, dual-purpose Indian banana cultivar.

Index Terms- banana ripeness classification, Nendra Bale, Random Forest, SVM, k-Nearest Neighbours

I. INTRODUCTION

Fruits are a cornerstone of a balanced, nutritious diet, and bananas rank among the most widely consumed fruits worldwide owing to their potassium, vitamin, and antioxidant content [1]. India is among the largest producers, contributing close to a quarter of global output [2], which underscores the crop's importance to both food security and the agricultural economy.

However, bananas are highly perishable climacteric fruits: rapid ripening after harvest drives substantial post-harvest losses during storage, transport, and distribution, directly affecting farmers, distributors, and exporters.

Fruit quality is largely determined by ripeness stage, which influences consumer acceptance, market value, and optimal consumption time. Ripeness is traditionally judged by manual inspection—visual grading and colour comparison—but these methods are subjective, labour-intensive, and inconsistent at scale. This has created strong demand for automated, reliable solutions for ripeness characterization across agricultural supply chains.

Among the many Indian banana cultivars, Nendra Bale (Nendran) is a dual-purpose, starch-rich variety widely used in both table consumption and cooking, particularly prominent in South Indian cuisine. Unlike commercially standardized, colour-dominant cultivars, Nendra Bale's ripening is governed substantially by its high starch and dietary fibre content, producing gradual, continuous texture transitions with comparatively muted colour change. This texture-dominant behaviour makes the variety a valuable contrasting case study.

Classical machine-learning models such as SVM, kNN, and RF are widely used for fruit classification because they are simple, interpretable, and effective on structured, feature-based inputs; SVM-based colour/texture approaches, for example, have shown promise for banana ripeness classification [4]. However, most prior work targets classification alone, rarely integrates shelf-life estimation, and depends on small or controlled datasets that limit generalization

[7]—a gap especially acute for texture-dominant, regionally significant varieties such as Nendra Bale, for which no dedicated classification study appears to exist.

Contributions. This paper does three things. First, we build a ripeness-labelled image set for Nendra Bale collected under everyday mobile-camera conditions, the first such structured dataset documented for this cultivar to the authors' knowledge. Second, we put three classical classifiers through the same test on both raw and augmented versions of the dataset. Third, we propose a probability-weighted rule that converts a predicted ripening stage into an estimated number of days remaining before spoilage.

II. RELATED WORK

Image-based, non-destructive ripeness assessment of bananas has been approached along two lines. The first applies classical machine learning to hand-crafted features—colour histograms, texture descriptors, and HOG—using classifiers such as SVM, kNN, and RF; these pipelines are interpretable and efficient on small-to-medium datasets typical of field studies. The second uses deep CNNs that learn features directly from pixels, achieving strong accuracy at the cost of larger data and compute requirements. Table I positions representative studies against the present work.

TABLE I. Representative Prior Work and the Gap Addressed

Study	Focus	Gap Addressed
Hou et al. [4]	Peel colour + SVM	Colour-dominant variety only
Mazen & Nashat [13]	ANN-based ripeness classification	No cross-variety or texture-dominant comparison
CNN studies [5]	Deep CNNs, larger data	No texture-dominant cultivar
Gulzar [6]	MobileNetV2 transfer learning	Generic fruits; no shelf life
Martínez-Mora et al. [14]	Dual CNN/VGG vs classical ML	Controlled lab setting; no texture-dominant case
Baglat et al. [7]	Systematic review of 35 studies	Scarce datasets; four-stage limitation noted
Shinga et al. [15]	Firmness prediction, coated bananas	Firmness as proxy; no image-based staging
This work	HOG+SVM/kNN/RF, raw vs aug.	First texture-dominant Indian cultivar dataset + shelf-life mapping

Three gaps recur across this literature: the scarcity of datasets for texture-dominant, regionally significant Indian cultivars such as Nendra Bale; the rarity of studies that explicitly compare raw versus augmented data for a cultivar whose ripening transitions are texture-led rather than colour-led; and the near-total absence of work that places a texture-dominant cultivar alongside colour-dominant ones under an identical experimental protocol so that the effect of cultivar biology on model choice can be isolated. This study targets all three gaps for Nendra Bale, building on the companion studies for Robusta and Nanjangud Rasabale conducted under the same broader research programme.

III. MATERIALS AND METHODS

A. Dataset Development

A structured image dataset of Nendra Bale was self-collected for this study, alongside Robusta and Nanjangud Rasabale datasets developed in the broader research programme. Unlike those colour/aroma-dominant cultivars, Nendra Bale's ripening is comparatively more texture-driven, making it a valuable contrasting case.

B. Ripening-Stage Definition

Samples were categorised into four stages consistent with the protocol used for other cultivars in this

programme: Green (unripe), Mid-ripe (yellowish-green), All-Yellow/Ripe (fully yellow), and Overripe (brown spotting and softening).

C. Image Acquisition

Images were captured under indoor and outdoor settings, natural and shaded light, and diverse backgrounds. A mobile-phone camera (Apple iPhone 12; 4 mm focal length; no flash) was used, consistent with the protocol applied across other varieties in this study.

D. Dataset Size and Distribution

The raw dataset comprises 80 images distributed unevenly across the four stages, with Green and Mid-ripe under-represented relative to All-Yellow and Overripe. Augmentation (rotation, flipping, brightness/contrast adjustment, scaling, translation, background variation) enlarged the training set to 5,391 images. Table II summarises counts.

TABLE II. Nendra Bale Dataset Composition

Stage	Raw	Augmented
Green	10	491
Mid-ripe	14	980
All-Yellow	32	2,240
Overripe	24	1,680
Total	80	5,391

E. Pre-processing and Feature Extraction

Images were normalised, denoised, and encoded with Histogram of Oriented Gradients (HOG) descriptors, capturing edge, shape, and texture cues—particularly relevant for Nendra Bale given its texture-driven ripening.

F. Classification Models

SVM (RBF kernel): draws the widest margin between ripening-stage classes, expected to help where visual change is gradual and texture-led rather than sharply segmented [9].

kNN: labels a new image by similarity to nearby training examples, an intuitive benchmark dependent on how well same-stage images cluster in feature space.

Random Forest (RF): grows many decision trees and averages votes; its threshold-based splits can be a disadvantage when ripening transitions are smooth and continuous [10].

G. Shelf-Life Estimation Module

The predicted ripening stage is mapped to an expected remaining shelf life in days using a probabilistic rule consistent with other cultivars in this programme: Green 8–12 days, Mid-ripe 5–8 days, All-Yellow 2–5 days, Overripe 0–2 days. The expected remaining life $S = \sum p_i \cdot d_i$, where p_i is the predicted probability of stage i and d_i is the midpoint of its range.

H. Experimental Setup

Classical models were evaluated separately on the raw dataset (80 images, 17-image held-out test set) and the augmented dataset (5,391 images). Experiments were implemented in Python (scikit-learn, scikit-image); performance was measured using classification accuracy.

IV. RESULTS AND DISCUSSION

A. Raw Dataset Results

On the raw dataset, kNN achieved the highest accuracy (76.47%), ahead of SVM and Random Forest (70.59% each). With only 80 images and a 17-image test set, these results should be interpreted cautiously—a single misclassification shifts accuracy by ~5.9 points. Table III summarises results.

TABLE III. Accuracy — Raw Dataset

Model	Accuracy (%)
SVM (RBF)	70.59
kNN	76.47
Random Forest	70.59

B. Augmented Dataset Results

Once enlarged through augmentation, the ranking changes: SVM becomes strongest at 81.60%, ahead of kNN (78.62%) and Random Forest (74.53%). Table IV summarises results, and Fig. 1 visualises the raw-versus-augmented comparison across all three classifiers.

TABLE IV. Accuracy — Augmented Dataset

Model	Accuracy (%)
SVM (RBF)	81.60
kNN	78.62
Random Forest	74.53

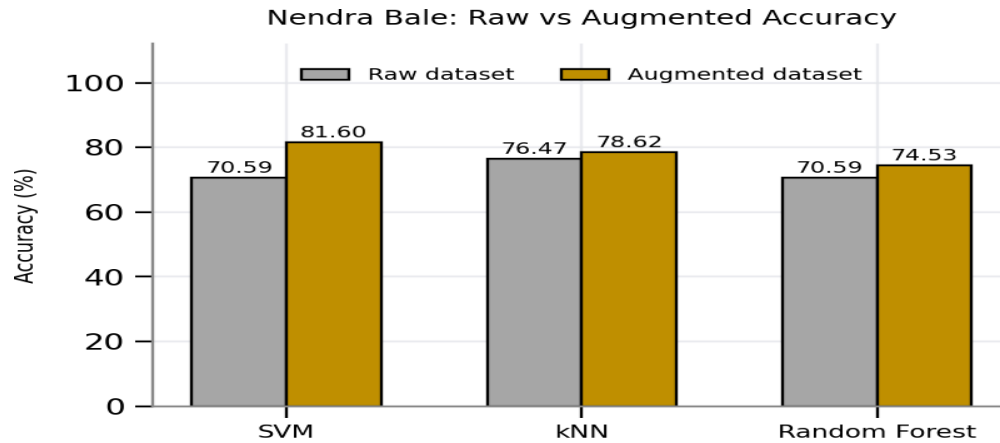


Fig. 1. Nendra Bale classical ML accuracy: raw vs augmented dataset.

This reversal mirrors the pattern seen for other cultivars in this programme: kNN's raw-dataset advantage appears to stem from the small, imbalanced sample rather than genuinely superior modelling. On the larger set, SVM's margin-based boundary generalises better to Nendra Bale's gradual texture transitions, while Random Forest's threshold-based splits remain poorly suited to this smooth progression. This pattern is consistent with the dataset-scale sensitivity documented for Robusta in the companion study, where Random Forest's raw-dataset lead (83.33%) similarly collapsed once the dataset was balanced (65.93%), reinforcing that apparent raw-dataset strength for tree-based ensembles should not be taken as a reliable indicator of true generalisation ability.

C. Why SVM Leads for Nendra Bale

SVM's superior performance (81.60%) stems from the texture-dominant ripening of Nendra Bale, governed by high starch and fibre content. Unlike colour-driven cultivars, visual change is gradual and continuous, with muted colour shift while unripe. SVM models non-linear boundaries effectively in this setting; kNN suffers from overlapping neighbourhoods, and RF's

threshold splits are unsuited to smooth texture evolution. Table V summarises the underlying nutrient–texture relationship.

TABLE V. Nutrient–Texture Relationship

Nutrient	Impact on Fruit	ML Feature Impact
High starch	Dense pulp, firmness	Strong gradient edges
Fibre	Fibrous texture	High HOG variance
Potassium	Cell turgor	Gradual texture change
Low sugars (unripe)	Less colour change	Texture > colour

D. Comparative Context Across the Research Programme

To place Nendra Bale's results in context, Fig. 2 compares the best classical accuracy (SVM, augmented dataset) achieved for all three Indian cultivars studied under this broader research

programme: Robusta (85.32%), Nanjangud Rasabale (88.60%), and Nendra Bale (81.60%).

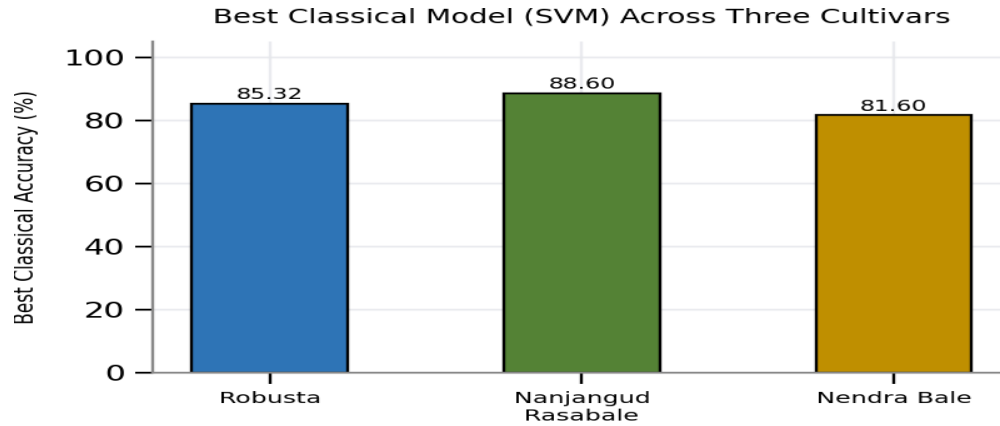


Fig. 2. Best classical model (SVM) accuracy across three Indian banana cultivars.

The lower absolute accuracy for Nendra Bale relative to Robusta and Nanjangud Rasabale is attributable to two compounding factors: the smaller, more imbalanced raw dataset (80 images, with as few as 10 in the Green class, compared with 202 for Robusta and 446 for Rasabale), and the texture-dominant, visually subtle ripening characteristics of the variety itself, which produce overlapping HOG feature distributions between adjacent stages even after augmentation. This does not indicate a deficiency in the modelling methodology; rather, it highlights that texture-dominant cultivars pose a genuinely harder classification problem than colour- or aroma-dominant ones when using the same hand-crafted feature representation.

Fig. 3 extends this comparison to all three classical models side by side. A consistent pattern emerges: SVM is the strongest or co-strongest classifier for every cultivar once the dataset is augmented, while Random Forest is consistently the weakest, regardless of whether the cultivar's ripening is colour-dominant (Robusta, Rasabale) or texture-dominant (Nendra Bale). This cross-cultivar consistency strengthens the case that SVM's margin-based decision boundary is better suited than Random Forest's threshold-based splits to the smooth, continuous nature of banana ripening in general—not merely to any one cultivar's specific ripening chemistry.

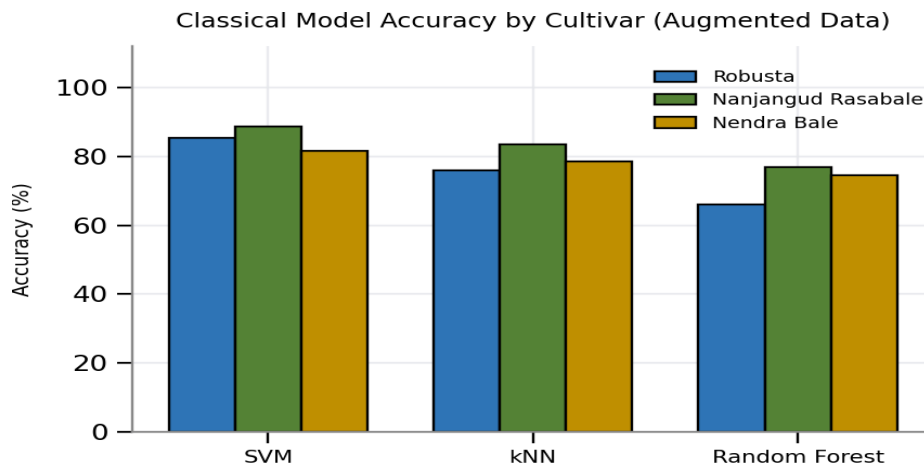


Fig. 3. Classical model accuracy by cultivar, all three models compared (augmented dataset).

At the same time, the magnitude of the gap between SVM and Random Forest differs meaningfully by cultivar: 19.4 percentage points for Robusta, 11.8 points for Rasabale, but only 7.1 points for Nendra Bale. This narrower gap suggests that when ripening is texture-driven rather than colour-driven, the relative advantage of a margin-based classifier over a tree-based ensemble is somewhat reduced—plausibly because HOG features, being texture-oriented by design, already partially compensate for Random Forest's weakness on smooth colour transitions. This

is a novel observation specific to comparing texture-dominant and colour-dominant cultivars under an identical feature-extraction pipeline, and motivates the colour–HOG fusion experiments proposed as future work.

Table VI consolidates the cross-cultivar interpretation, summarising the dominant ripening cue, best classifier, and SVM–Random Forest gap for each of the three varieties studied.

TABLE VI. Cross-Cultivar Summary (Augmented Dataset)

Cultivar	Dominant Cue	Best Model	SVM–RF Gap (pts)
Robusta	Colour + moderate texture	SVM (85.32%)	19.4
Nanjangud Rasabale	Colour + aroma	SVM (88.60%)	11.8
Nendra Bale	Texture (starch/fibre)	SVM (81.60%)	7.1

V. LIMITATIONS

Small, imbalanced raw dataset (80 images; 17-image test set) makes raw-dataset accuracies highly sensitive to individual misclassifications.

Non-uniform augmentation: the augmented set (5,391 images) reflects raw-stage imbalance rather than a fully class-balanced protocol.

Deep learning (Custom CNN, MobileNetV2), completed for Robusta and Rasabale, has not yet been reliably completed for Nendra Bale.

Shelf-life-in-days mapping is a literature-informed heuristic, not yet validated against measured spoilage data for this cultivar.

HOG captures texture but not colour explicitly; colour–HOG fusion may improve reliability. Data cover a single site; cross-site/season generalisation is untested.

VI. CONCLUSION

This study introduced a ripeness-labelled image dataset for Nendra Bale, a texture-dominant, dual-purpose Indian banana cultivar, and compared three classical ML models on raw and augmented data. On

the raw dataset, kNN led (76.47%) over SVM and RF (70.59% each); once augmented, SVM became the clear leader (81.60%), ahead of kNN (78.62%) and RF (74.53%), reflecting the gradual, continuous texture transitions of this cultivar's ripening. These findings mirror dataset-scale sensitivity observed for other cultivars in this research programme. Placed alongside the companion Robusta and Nanjangud Rasabale studies, the cross-cultivar comparison shows that SVM is the consistently strongest classical model regardless of whether ripening is colour-dominant or texture-dominant, though the margin by which it outperforms Random Forest narrows for texture-dominant cultivars—a novel, cultivar-specific insight that directly motivates the colour–HOG feature-fusion work proposed below.

Future work will complete CNN/MobileNetV2 experiments for Nendra Bale, validate shelf-life predictions against measured spoilage data, benchmark colour–HOG fusion, expand and rebalance the raw dataset, and report full per-class precision/recall/F1 metrics.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, for supporting this research.

VII. DECLARATIONS

Funding: The authors received no specific funding for this work.

Competing Interests: The authors declare that they have no competing interests.

Data Availability: The Nendra Bale image dataset generated and analysed in this study is available from the corresponding author on reasonable request.

Ethics Approval: This study did not involve human participants or animals; ethics approval was not required.

Author Contributions: Manasa T. N. conceived the study, collected the dataset, implemented the models, and drafted the manuscript. M. P. Pushpalatha supervised the work, provided methodological guidance, and reviewed and edited the manuscript. Both authors read and approved the final manuscript.

Declaration of Generative AI Use: During preparation of this manuscript, the authors used an AI-based writing assistant (Microsoft Copilot) to assist with formatting, language editing, and structuring of the text consistent with the IRE Journals template. The tool was not used to generate, analyze, or interpret the research data, results, or scientific conclusions, which are entirely the authors' own work. The authors reviewed and edited all AI-assisted content and take full responsibility for the content of this publication.

REFERENCES

- [1] FAO, FAOSTAT — Crops and Livestock Products: Bananas. Rome: Food and Agriculture Organization of the United Nations, 2024.
- [2] National Horticulture Board, Indian Horticulture Database — Banana Production Statistics. Ministry of Agriculture, Govt. of India, 2023.
- [3] E. Almeyda and W. Ipanaqué, “Recent developments in post-harvest handling and shelf-life assessment of bananas: a review,” *Sci. Hortic.*, vol. 305, pp. 111–123, 2022.
- [4] J. Hou, Y. Hu, L. Hou, K. Guo, and T. Satake, “Classification of ripening stages of bananas based on support vector machine,” *Int. J. Agric. Biol. Eng.*, vol. 8, no. 6, pp. 99–103, 2015.
- [5] N. Saranya, K. Srinivasan, and S. K. Pravin Kumar, “Banana ripeness stage identification using deep learning,” *J. Ambient Intell. Humaniz. Comput.*, vol. 13, pp. 4033–4039, 2022.
- [6] Y. Gulzar, “Fruit image classification model based on MobileNetV2 with deep transfer learning,” *Sustainability*, vol. 15, no. 3, art. 1906, 2023.
- [7] P. Baglat, A. Hayat, F. Mendonça, A. Gupta, S. S. Mostafa, and F. Morgado-Dias, “Non-destructive banana ripeness detection using shallow and deep learning: a systematic review,” *Sensors*, vol. 23, no. 2, art. 738, 2023.
- [8] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE CVPR*, 2005, pp. 886–893.
- [9] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: inverted residuals and linear bottlenecks,” in *IEEE CVPR*, 2018, pp. 4510–4520.
- [12] F. Pedregosa et al., “Scikit-learn: machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [13] F. Mazen and A. Nashat, “Ripeness classification of bananas using an artificial neural network,” *Arab. J. Sci. Eng.*, vol. 44, pp. 6901–6910, 2019.
- [14] O. Martínez-Mora, O. Capuñay-Uceda, L. Caucha-Morales, R. Sánchez-Ancajima, I. Ramírez-Morales, S. Córdova-Márquez, and F. Cuenca-Mayorga, “Artificial vision-based dual CNN classification of banana ripeness and quality attributes using RGB images,” *Processes*, vol. 13, no. 4, art. 1142, 2025.
- [15] M. Shinga, Y. Silué, and O. Fawole, “Machine learning-based prediction of firmness in coated

bananas under retail conditions,” *Postharvest Biol. Technol.*, vol. 220, art. 113276, 2025.

- [16] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.