International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 08 August-2026 | Impact Factor: 3.5

Robust Object Recognition Using NIG-SHAPE Hybrid Features and Stacking on COIL-100

Amar Chand[1], Ravinder Singh Madhan[2]

[1] Research Scholar, Department of Computer Science & Engineering, IEC University, Baddi, Solan (H.P.), India
[2] Associate Professor, Department of Computer Science & Engineering, IEC University, Baddi, Solan (H.P.), India

Corresponding Author Email: amar.chand@csu.co.in

How to Cite this Article:

Chand, A. (2026). Robust Object Recognition Using NIG-SHAPE Hybrid Features and Stacking on COIL-100. International Journal of Creative and Open Research in Engineering and Management, 02(8), 1-9. https://doi.org/10.55041/ijcope.v2i8.199

https://doi.org/10.55041/ijcope.v2i8.199

Abstract—

Object recognition on controlled multi-view benchmarks can produce deceptively high clean accuracy, making it difficult to distinguish a genuinely informative representation from a model that mainly exploits stable appearance. This study develops and evaluates an interpretable recognition pipeline in which geometric information remains explicit while texture, color, and classifier diversity provide complementary evidence. The proposed Normalized Invariant Geometry descriptor (NIG-SHAPE) combines log-transformed moment invariants, normalized Fourier contour magnitudes, radial occupancy measurements, and compact contour geometry statistics. The 45-dimensional shape block is fused with a 26-dimensional Local Binary Pattern histogram and 96 RGB/HSV histogram features, producing a standardized 167-dimensional representation. Classification is performed by a Hybrid Feature Stacking Ensemble that integrates Random Forest, Linear Support Vector Machine, K-Nearest Neighbour, and Histogram Gradient Boosting predictions through a Logistic Regression meta-learner. Experiments use the corrected 7,200-image COIL-100 pipeline with a stratified 80:20 split and deterministic stress augmentation for the final model. NIG-SHAPE improves shape-only accuracy from 0.6424 for Hu moments to 0.8861, while the full hybrid representation reaches the clean-set ceiling. The robustness-augmented ensemble obtains 1.0000 clean accuracy and retains 0.9972 under occlusion, 0.9979 under lighting shift, and 0.9993 under clutter. The results show that the main value of the framework lies not in a clean-score superiority claim, but in a transparent combination of stronger geometry, complementary cues, classifier diversity, and disturbance-aware evaluation.

Keywords— object recognition; shape descriptor; feature fusion; COIL-100; stacking ensemble; robustness

I. INTRODUCTION

Reliable object recognition requires more than assigning labels to clean images. A practical system must retain identity when viewpoint, visible boundary, illumination, or background context changes, and it should also provide enough structure for the researcher to explain why a decision is plausible. These requirements remain relevant even as deep neural models dominate contemporary computer vision. Learned representations are powerful, but for controlled datasets and research questions centered on geometry, a transparent handcrafted pipeline can still be valuable because individual descriptor components can be isolated, measured, and tested under controlled perturbation. Shape-based studies continue to explore contour meshes, local shape points, frequency representations, and invariant geometry for this reason [1]–[4].

Shape is especially attractive for multi-view recognition because it encodes the external and internal organization of an object rather than only its surface appearance. However, a single global descriptor can

lose important detail. Moment invariants are compact and convenient, yet two objects with comparable gross geometry may remain difficult to separate if their local contour structure differs. Conversely, contour descriptors can become unstable when segmentation is imperfect or part of the boundary is hidden. Recent work on shape matching and contour-oriented recognition illustrates both the usefulness of geometric structure and the need for descriptors that combine more than one geometric view [2], [3], [5].

Texture and color supply information that shape cannot provide. Local Binary Pattern (LBP) variants have remained useful for compact local texture description, while color-texture methods show that local intensity structure and chromatic distribution can complement one another [6]–[8]. Feature-fusion studies similarly demonstrate that combining heterogeneous cues can improve discrimination when no single family is sufficient [9]–[11]. The difficulty is that each cue has its own failure mode: color is highly effective under stable acquisition but can change with illumination; texture may be altered by blur or local noise; and shape depends on a reliable foreground mask. A strong recognition framework therefore needs both cue diversity and an evaluation protocol that exposes these weaknesses rather than hiding them behind a single clean score.

Classifier choice introduces a second source of variation. Random Forest, Support Vector Machine, K-Nearest Neighbour, and boosting methods impose different decision biases, so the classifier that works best for one representation is not automatically best for another [12], [13]. Stacking provides a natural way to exploit this diversity by learning how to combine predictions from several base learners. Stacking has been applied successfully in multiclass recognition and hybrid image-classification settings [14], [15], suggesting that model-level fusion can complement feature-level fusion when the base learners capture different structures in the feature space.

The present study addresses four linked gaps. First, a moment-only shape baseline is too coarse for a 100-object multi-view problem. Second, single-cue systems cannot fully exploit the stable color and texture evidence available in COIL-100 while remaining prepared for appearance disturbance. Third, clean benchmark accuracy can become saturated, which makes a superiority claim based only on the clean split scientifically weak. Fourth, a single classifier does not reveal whether performance depends on one decision bias. The proposed solution is a modular pipeline centered on a new Normalized Invariant Geometry shape descriptor (NIG-SHAPE), fused with LBP and RGB/HSV histograms, and classified by a Hybrid Feature Stacking Ensemble (HFS-Ensemble).

The study is designed to answer three practical questions. How much does a multi-part geometric descriptor improve over Hu moments when shape is evaluated alone? How do shape, texture, and color behave when tested separately and in combination? Finally, can a robustness-aware stacking model preserve the clean-set ceiling while remaining stable under controlled occlusion, lighting shift, and clutter? The paper therefore emphasizes ablation, computational behavior, and stressed evaluation in addition to clean accuracy. An earlier version of the overall framework was reported by the authors [20]; the present manuscript reorganizes the thesis evidence around descriptor contribution, clean-set saturation, and robustness interpretation rather than reproducing that article.

II. LITERATURE REVIEW

A. Shape Representation and Invariance

Recent shape-based recognition research consistently shows that useful descriptors must suppress nuisance variation while retaining class-specific structure. Circular mesh and margin descriptors represent object boundaries through a richer spatial organization than a single global statistic [1]. Shape-interest-point methods similarly use local geometric evidence to improve recognition when different portions of a contour carry unequal discriminative value [2]. Frequency-domain representations provide another view: weighted Fourier and wavelet-like descriptors transform contour variation into compact coefficients that can be normalized for geometric invariance [3]. These directions support an important design principle for the present work: a robust shape vector should not depend on one mathematical family alone.

The limitation of single-family geometry becomes clearer when objects share similar silhouettes. Global moments summarize broad spatial structure, but they do not explicitly encode how a contour changes around the object, how mass is distributed directionally around the centroid, or how compact and convex the boundary is. Combining algebraic moments with contour-

frequency and geometry statistics can therefore preserve both global and boundary-specific information. Contemporary shape descriptors continue to exploit this combination of local, regional, and invariant representations [2]–[5]. NIG-SHAPE follows this line of reasoning while keeping each component separately interpretable.

B. Texture, Color, and Feature Fusion

Texture is often used when boundary shape alone cannot separate objects with similar outlines. LBP-based descriptors remain attractive because they encode local neighborhood transitions without requiring a large learned model. Color-texture LBP has produced effective image-retrieval representations [6], and circular or directional LBP variants have been proposed to improve stability under rotation and local pattern change [7], [8]. These studies motivate the use of a compact LBP histogram as supporting evidence rather than as the primary representation.

Color provides a different type of evidence. On controlled datasets, stable object appearance can become extremely discriminative; however, the same strength creates a methodological risk because a classifier may solve the benchmark mainly from color and learn little about geometry. Hybrid feature research addresses this issue by combining appearance and structural cues. Color-and-shape fusion [9], learned and handcrafted feature fusion [10], and LBP-color-histogram fusion [11] all illustrate the broader value of heterogeneous representation. The present study adopts early fusion so that the contributions of shape, texture, and color can be compared directly before they enter the classifier.

C. Classifier Diversity, Stacking, and Robustness

Classical classifiers respond differently to the same feature space. Tree ensembles capture nonlinear interactions, linear SVM emphasizes margins in standardized space, and KNN preserves local neighborhood relations. Comparative studies report meaningful variation among these models, reinforcing the view that no single learner should be assumed optimal a priori [12], [13]. Stacking extends this argument by converting base-model outputs into a second-level representation. Applications in celestial-object classification and traffic-sign recognition show that stacking can improve multiclass decision making when component models provide complementary errors [14], [15].

Robustness is equally important because clean accuracy does not reveal how a representation behaves when part of the evidence is disturbed. Occlusion-aware recognition research has shown that partial visibility changes the reliability of feature evidence and often requires explicit adaptation or search strategies [16], [17]. Work on recognition in clutter likewise emphasizes the need to distinguish the object from competing background structure [18], while illumination-oriented studies show that image quality and acquisition conditions can strongly affect object recognition [19]. These findings justify a separate stress protocol rather than treating perturbation as an incidental observation.

The gap emerging from these literature streams is not the absence of shape descriptors, texture features, color histograms, or ensemble methods individually. The gap is the absence, within the scope of this study, of a transparent pipeline that strengthens shape first, exposes the contribution of each feature family, combines diverse classical learners, and then tests the final model under multiple controlled disturbances. This gap is particularly relevant to COIL-100 because its controlled background and appearance can push several models to perfect clean accuracy. Under such saturation, the scientific contribution must be argued through representation quality, stress behavior, interpretability, and reproducibility rather than through a single headline score.

III. METHODOLOGY

A. Dataset and Experimental Protocol

The experiments use the corrected COIL-100 pipeline containing 100 object classes and 72 rotational views per object, giving 7,200 canonical images. Every input is resized to 128 × 128 pixels. A stratified 80:20 split produces 5,760 training images and 1,440 held-out test images, preserving approximately equal support across the 100 classes. Baseline models are trained only on the clean training set. The final robustness-aware model uses one deterministic stress variant for each clean training image, producing 11,520 training samples while leaving the held-out test images untouched during training.

Table I: Experimental configuration of the proposed framework.

Parameter	Value
Dataset	COIL-100 corrected pipeline
Classes / images	100 / 7,200
Views per object	72
Input resolution	128 × 128 pixels
Clean split	5,760 train / 1,440 test
Augmented train set	11,520 images
Feature dimensions	45 shape + 26 LBP + 96 color = 167
Stress tests	Occlusion, lighting shift, clutter
Base learners	RF, Linear SVM, KNN, Hist. Gradient Boosting
Meta-learner	Logistic Regression

The experimental design deliberately separates baseline discrimination from robustness-aware training. This avoids giving the final ensemble an unfair advantage in the clean baseline comparison and makes it possible to attribute performance gains to the feature families before stress augmentation is introduced. Accuracy, macro-F1, and weighted-F1 are reported for all principal evaluations. Macro-F1 is important in a 100-class task because it weights every class equally, while weighted-F1 confirms that the result is not being driven by unequal class support.

B. Preprocessing and Foreground Isolation

Preprocessing is treated as part of the descriptor rather than as a cosmetic preparation stage. Each image is retained in RGB for color analysis, converted to HSV for complementary color representation, and converted to grayscale for masking and LBP extraction. A Gaussian blur suppresses small pixel-level variation before Otsu thresholding separates the foreground from the controlled background. Morphological opening removes isolated artifacts, and closing fills small gaps in the object mask. The largest connected contour is then selected as the principal object boundary.

This sequence is especially important for shape analysis because an unstable mask directly changes moment values, contour frequencies, radial measurements, and compactness statistics. By using one deterministic preprocessing path for all models, the experiment keeps the comparison focused on descriptor and classifier differences rather than on inconsistent object extraction. The approach is well suited to COIL-100, although the threshold-based foreground assumption is later treated as a limitation for natural-scene deployment.

C. NIG-SHAPE Descriptor

NIG-SHAPE contains four complementary geometric blocks. The first block uses log-transformed invariant moments. If phi_k denotes the k-th moment invariant, the log mapping compresses its large dynamic range and preserves sign:

$$h_k = -sign(phi_k)\ log10(|phi_k| + epsilon) \quad (1)$$

The second block represents the boundary in the frequency domain. Contour points are centered at the contour centroid and normalized by the mean radius before a discrete Fourier transform is applied. Magnitudes, rather than complex phases, are retained to reduce rotation sensitivity, and the coefficients are normalized by the first nonzero component to reduce scale influence. In simplified form, the normalized coefficient is

$$f_m = |F_m| / (|F_1| + epsilon) \quad (2)$$

The third block measures radial occupancy. Foreground pixels are grouped into angular bins around the object centroid, the mean radius in each bin is computed, and the resulting directional pattern is normalized by the overall mean radius. This captures how object mass extends in different directions, providing information that neither global moments nor contour frequency alone represents. The fourth block contains explicit contour geometry measurements, including area ratio, normalized perimeter, circularity, extent, solidity, aspect ratio, equivalent diameter, eccentricity, edge density, and a fill-convexity term. Together the four blocks form a 45-dimensional shape vector.

The novelty of this construction is organizational rather than dependent on an opaque learned embedding. Every component has a geometric interpretation: moments summarize global region structure; Fourier magnitudes describe global boundary variation; radial occupancy describes directional mass distribution; and contour statistics describe compactness and shape regularity. Because all components are computed from the same foreground mask and normalized before fusion, the descriptor remains compact enough for classical machine learning while offering more information than a moment-only baseline.

D. Texture, Color, and Early Fusion

Texture is represented by a 26-dimensional LBP histogram. For each grayscale pixel, the eight neighboring intensities are compared with the center

pixel to produce a local binary code. The image-level histogram summarizes the frequency of these local transitions, giving a compact description of repeated micro-structure. Color is represented through normalized histograms from both RGB and HSV spaces. RGB preserves direct channel intensity, whereas HSV separates hue and saturation from value and therefore provides an alternative view of chromatic appearance. The combined color block has 96 dimensions.

The three feature families are concatenated through early fusion. If s_i is the 45-dimensional shape vector, t_i is the 26-dimensional LBP vector, and c_i is the 96-dimensional color vector, then

$$x_i = [s_i, t_i, c_i] \quad (3)$$

The final vector contains 167 features. Because the three blocks have different numeric ranges, standardization is performed using means and standard deviations estimated from the training set only. This prevents the larger color block or high-magnitude geometric variables from dominating distance- and margin-based classifiers. The same scaler is then applied to the held-out test data, preserving the separation between training and evaluation.

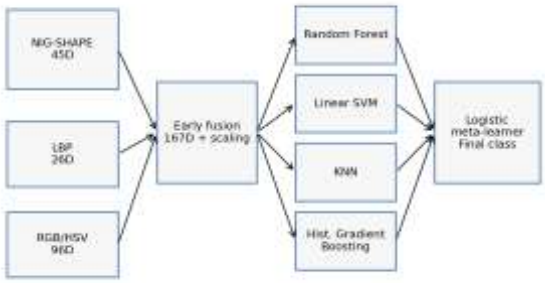

Figure 1: NIG-SHAPE, feature fusion, and HFS-Ensemble architecture.

E. Hybrid Feature Stacking Ensemble

The HFS-Ensemble uses Random Forest, Linear SVM, KNN, and Histogram Gradient Boosting as base learners. The selection is intentionally heterogeneous. Random Forest models nonlinear feature interactions through an ensemble of decision trees; Linear SVM tests whether the standardized hybrid representation is close to linearly separable; KNN preserves local neighborhood structure; and Histogram Gradient Boosting captures nonlinear residual patterns through additive refinement. Their outputs are collected into a second-level vector rather than averaged directly.

$$z_i = [B_1(xhat_i), B_2(xhat_i), B_3(xhat_i), B_4(xhat_i)] \quad (4)$$

Logistic Regression acts as the meta-learner and estimates the final multiclass decision from the combined base outputs:

$$yhat_i = arg\ max_c\ P(c \mid z_i) \quad (5)$$

A simple meta-learner is preferred because it keeps the decision layer interpretable and avoids moving all modeling complexity into a second opaque model. The baseline HFS-Ensemble is trained on clean data. The final augmented model is trained on the 5,760 clean training images plus 5,760 deterministic stress variants. The test set is evaluated in four conditions: clean, occluded, lighting-shifted, and cluttered. The perturbations are synthetic and reproducible, providing a controlled first test of disturbance tolerance without claiming to reproduce the full variability of natural scenes.

F. Evaluation Metrics and Analysis Strategy

The analysis is organized in three layers. First, feature-family baselines isolate the contribution of Hu moments, NIG-SHAPE, LBP, color histograms, and fused descriptors. Second, multiple classifiers are compared on the hybrid representation to determine whether the clean-set ceiling depends on a single learning algorithm. Third, the robustness-augmented HFS-Ensemble is evaluated under three disturbances. Training and inference times are retained where available so that computational cost can be discussed separately from recognition quality. This structure is necessary because perfect clean accuracy on a controlled benchmark is not, by itself, sufficient evidence of a better recognition framework.

IV. RESULTS AND DISCUSSION

A. Clean Baselines and the Clean-Set Ceiling

The clean experiments reveal two different stories at the same time. At the feature-family level, the representation matters substantially: Hu moments alone are weak, NIG-SHAPE is much stronger, and LBP adds another strong source of discrimination. At the full-system level, however, COIL-100 is sufficiently controlled that several hybrid baselines reach 1.0000 accuracy and 1.0000 macro-F1. This saturation means that the final ensemble cannot be defended by claiming a unique clean-score advantage. Instead, the more informative evidence lies in the

progression from weak to strong descriptors and in the stability of the final model under disturbance.

Table II: Selected clean-test baseline and proposed-model results.

Model	Accuracy	Macro-F1	Train (s)
Hu moments + RF	0.6424	0.6312	4.3361
NIG-SHAPE + RF	0.8861	0.8820	8.9899
LBP + RF	0.9181	0.9154	4.7490
Color hist. + RF	1.0000	1.0000	8.3582
Hu+LBP+Color + RF	1.0000	1.0000	10.2626
NIG+LBP+Color + ET	1.0000	1.0000	2.5300
NIG+LBP+Color + SVM	1.0000	1.0000	253.6434
NIG-Hybrid HFS	1.0000	1.0000	546.3599
Augmented NIG-Hybrid HFS	1.0000	1.0000	—

The most direct descriptor result is the improvement from 0.6424 accuracy for Hu moments to 0.8861 for NIG-SHAPE using the same Random Forest classifier. The absolute gain of 0.2437 is large enough to show that the additional contour, radial, and geometry information is not redundant. Macro-F1 rises in parallel from 0.6312 to 0.8820, indicating that the gain is distributed across classes rather than concentrated in a small subset. LBP-only reaches 0.9181 accuracy, confirming that local texture is highly informative for these objects. This ordering also clarifies the role of NIG-SHAPE: it is not expected to dominate all appearance cues on a controlled color dataset; its purpose is to provide a stronger and more interpretable geometric foundation for fusion.

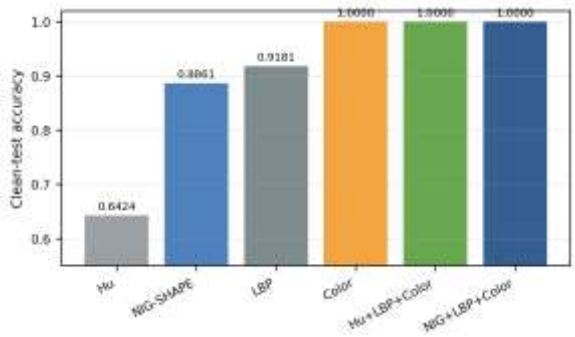

Figure 2: Best clean-test accuracy obtained by major feature families.

Color histograms alone reach 1.0000 with Random Forest. This is an important result precisely because it limits the strength of any claim based on clean accuracy. COIL-100 contains stable object colors and a controlled background, so appearance can be unusually discriminative. The correct interpretation is therefore not that shape is unnecessary, but that clean COIL-100 classification is a saturated test for several strong feature combinations. A shape-centered research contribution must demonstrate why its geometry is useful through ablation, perturbation behavior, and transparency rather than by ignoring a perfect color baseline.

The hybrid models confirm that the fused 167-dimensional representation is easy for several classifier families to separate. Random Forest, Extra Trees, and Linear SVM all reach the clean ceiling on either the Hu-based or NIG-based hybrid vector. KNN is slightly lower: Hu+LBP+Color with KNN reaches 0.9958, while NIG-SHAPE+LBP+Color with KNN reaches 0.9882. The KNN result is useful because it shows that simply adding a stronger shape block does not guarantee improvement for every learner. Local distance structure can respond differently to the geometry of a high-dimensional fused space, which is one reason to compare classifier biases rather than reporting only the strongest model.

B. Feature Complementarity and Descriptor Contribution

The ablation pattern supports a complementary-cue interpretation. Hu moments capture global geometry but lose too much detail for 100 object identities. NIG-SHAPE restores a substantial amount of discrimination by adding contour-frequency, directional, and compactness information. LBP contributes local surface structure that may remain distinctive even when two objects share a similar outline. Color provides highly discriminative appearance in the clean acquisition setting. When these cues are fused, multiple classifiers reach perfect clean separation, showing that the combined vector contains redundant but useful evidence rather than relying on a single fragile dimension.

This behavior is consistent with the broader literature on heterogeneous fusion [9]–[11]. The value of fusion is not that each added block must increase a clean score that is already saturated. Rather, different cues can remain useful under different disturbances. Shape offers structural evidence when color changes; color and texture can compensate when a contour is partly hidden; and classifier diversity can stabilize decisions when local and global structures favor different boundaries. The robustness experiment is therefore the more meaningful test of whether the hybrid design has practical value.

C. Robustness Under Occlusion, Lighting Shift, and Clutter

The final robustness-augmented HFS-Ensemble preserves 1.0000 accuracy, macro-F1, and weighted-F1 on the clean held-out split. Under synthetic occlusion, accuracy decreases only to 0.9972 and macro-F1 to 0.9973. Under lighting shift, both accuracy and macro-F1 are 0.9979. Under clutter, both metrics remain at 0.9993. The absolute drops from clean accuracy are therefore 0.0028, 0.0021, and 0.0007 respectively.

Table III: Robustness of the final augmented HFS-Ensemble.

Condition	Accuracy	Macro-F1	Drop
Clean	1.0000	1.0000	0.0000
Occlusion	0.9972	0.9973	0.0028
Lighting shift	0.9979	0.9979	0.0021
Clutter	0.9993	0.9993	0.0007

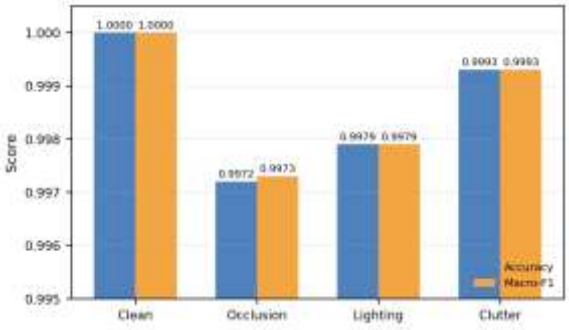

Figure 3: Accuracy and macro-F1 under clean and stressed conditions.

Occlusion produces the largest degradation, which is expected because it directly removes shape and surface evidence from the object. Even so, the reduction is small. The result suggests that the final classifier is not dependent on any single visible region. This interpretation agrees with the motivation of occlusion-aware recognition studies, where partial visibility changes the reliability of specific cues [16], [17]. In the present handcrafted framework, robustness is achieved not through an explicit occlusion detector but through redundant feature evidence and stress-aware training.

Lighting shift affects the appearance cues more directly. The 0.9979 accuracy indicates that the combined representation does not collapse when brightness is modified, even though color is exceptionally strong on the clean dataset. Shape measurements, HSV representation, standardization, and augmented training all provide routes for preserving identity. The result should still be interpreted within the limits of synthetic disturbance; real illumination changes can involve shadows, spectral shifts, specular highlights, and camera-response effects that are not reproduced by a simple brightness transformation [19].

Clutter causes the smallest decline, with accuracy of 0.9993. Because foreground extraction is based on thresholding and morphology, this result indicates that the tested clutter perturbation does not usually defeat the principal-contour selection step. Recognition in genuinely cluttered scenes is a harder problem because competing objects and textured backgrounds can corrupt segmentation more severely [17], [18]. The present result therefore demonstrates stability to the implemented stress condition rather than a general solution to arbitrary scene clutter.

D. Computational Behavior

Perfect clean accuracy is achieved by models with very different computational costs. Extra Trees on the NIG-SHAPE hybrid vector trains in approximately 2.53 s, whereas Linear SVM on the same hybrid representation requires 253.64 s. The unaugmented HFS-Ensemble requires 546.36 s and approximately 1.41 s for inference on the test set. This difference shows that stacking should not be chosen merely because it can reach a score that simpler models already achieve on the clean benchmark.

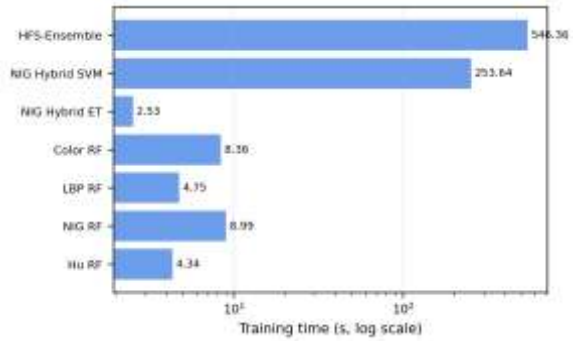

Figure 4: Training-time differences for representative models.

The ensemble cost is justified only if the research objective includes robustness, model diversity, and a unified evaluation framework. For applications where inputs remain as controlled as the clean COIL-100 images, a tree ensemble or even a color-histogram Random Forest may be a more efficient engineering choice. For research that emphasizes a shape-centered, disturbance-aware pipeline, the additional training cost can be acceptable because training is performed offline and the resulting system preserves clean performance while remaining strong across the three stressed test conditions. This distinction between an engineering optimum and a research contribution is important and prevents the final model from being oversold.

E. Comparison with Prior Work and Methodological Implications

The experimental evidence aligns with several themes in previous research without supporting a direct numerical ranking across unrelated datasets. Contour-rich shape descriptors remain useful because multiple geometric views capture information beyond global moments [1]–[5]. LBP and color-texture methods remain effective complementary representations [6]–[8]. Feature-fusion studies support the combination of structural and appearance cues [9]–[11], while stacking research supports the use of diverse base learners for multiclass decisions [14], [15]. The present work combines these directions in a deliberately interpretable pipeline and adds a stress protocol that makes the clean-set ceiling easier to interpret.

A central methodological finding is that benchmark saturation changes what should count as evidence. If only the best clean score were reported, the color baseline, several single-classifier hybrids, and the proposed ensemble would all appear equivalent at 1.0000. The ablation results reveal that they are not equivalent representations: Hu moments are weak, NIG-SHAPE is much stronger, LBP is stronger still, and color dominates clean appearance. Robustness then adds a second axis of evidence by showing that the augmented hybrid ensemble remains near the clean ceiling when three different types of disturbance are applied. The contribution is therefore the chain of evidence from descriptor design to perturbation-aware evaluation, not the isolated value 1.0000.

The close agreement between accuracy, macro-F1, and weighted-F1 also matters. With approximately balanced class support in the stratified split, the near-identical metrics indicate that performance is not being produced by a few easy classes while other classes fail. The thesis reports a diagonal clean confusion matrix and class-wise precision, recall, and F1 of 1.0 for the final clean split. In a saturated benchmark this does not prove generalization to natural scenes, but it does confirm that the held-out clean evaluation is internally consistent across all 100 object identities.

F. Limitations

Four limitations define the scope of the conclusions. First, COIL-100 is controlled, so its stable background and color distributions are not representative of natural scenes. Second, foreground extraction relies on thresholding and morphology; more complex backgrounds may require learned segmentation, saliency, or instance masks. Third, the stress conditions are synthetic and should be treated as diagnostic perturbations rather than substitutes for real occlusion, illumination variability, or clutter. Fourth, the stacking model is computationally more expensive than several clean-set baselines. These limitations do not invalidate the results, but they restrict the claim to an interpretable, high-performing framework on the corrected COIL-100 protocol with demonstrated robustness to the specified perturbations.

Future validation should therefore move in two directions. The first is harder data: natural-scene object datasets, more realistic segmentation, compound perturbations, blur, scale distortion, and discontinuous viewpoint change. The second is stronger comparison: lightweight convolutional or transformer models should be evaluated under the same train-test split and the same perturbation protocol so that accuracy, robustness, training cost, and interpretability can be compared on equal terms. Such experiments would reveal whether the handcrafted hybrid representation remains competitive when the controlled appearance advantage of COIL-100 is reduced.

V. CONCLUSION

This study presents a shape-centered but deliberately hybrid object-recognition framework for the corrected COIL-100 pipeline. The proposed NIG-SHAPE descriptor combines moment invariants, normalized Fourier contour magnitudes, radial occupancy measurements, and explicit contour geometry in a 45-dimensional vector. Its value is demonstrated by a shape-only accuracy increase from 0.6424 with Hu moments to 0.8861 under the same Random Forest classifier. The descriptor is then fused with a 26-dimensional LBP texture histogram and 96 RGB/HSV color features to form a standardized 167-dimensional representation. A stacking ensemble combines Random Forest, Linear SVM, KNN, and Histogram Gradient Boosting through a Logistic Regression meta-learner.

The experiments also show why clean accuracy must be interpreted carefully. Color-only and several hybrid baselines reach 1.0000 on clean COIL-100, so the final model is not presented as the only method capable of perfect clean classification. Its stronger research value is the complete evidence chain: an improved geometric descriptor, explicit feature-family ablation, classifier

diversity, controlled stress augmentation, and robustness evaluation. The augmented HFS-Ensemble preserves 1.0000 clean accuracy and records 0.9972 under occlusion, 0.9979 under lighting shift, and 0.9993 under clutter. These results support the conclusion that complementary geometry, texture, color, and model diversity can produce a transparent and highly stable recognition system in a controlled multi-view setting.

The next stage should test the framework beyond controlled backgrounds and replace threshold-based foreground extraction with stronger segmentation while preserving the interpretable core of NIG-SHAPE. Direct comparison with lightweight deep models under identical robustness tests would further clarify the trade-off between transparency, computational cost, and generalization. The present evidence nevertheless establishes that carefully engineered shape information remains useful when it is strengthened, fused with complementary cues, and evaluated with more than a single clean benchmark score.

ACKNOWLEDGMENT

The authors acknowledge the academic support of the Department of Computer Science & Engineering, IEC University, and the open scientific community whose datasets, software tools, and published research enabled the experimental study.

REFERENCES

[1] M. G., S. Elizabeth, and S. Mathew Koshy, “Circular mesh-based shape and margin descriptor for object detection,” Pattern Recognition, vol. 84, pp. 97–111, 2018, doi: 10.1016/j.patcog.2018.07.004.

[2] L. Zhang and J. Pu, “Object recognition based on shape interest points descriptor,” Electronics Letters, vol. 60, no. 9, 2024, doi: 10.1049/ell2.13198.

[3] Y. Zheng, B. Guo, C. Li, and Y. Yan, “A Weighted Fourier and Wavelet-Like Shape Descriptor Based on IDSC for Object Recognition,” Symmetry, vol. 11, no. 5, p. 693, 2019, doi: 10.3390/sym11050693.

[4] E. E. Çakı and C. O. Gökçe, “A Novel Shape Descriptor for Object Recognition,” International Journal of Computational and Experimental Science and Engineering, vol. 9, no. 1, pp. 1–5, 2023, doi: 10.22399/ijcesen.1202300.

[5] M. Lazarou, B. Li, and T. Stathaki, “A novel shape matching descriptor for real-time static hand gesture recognition,” Computer Vision and Image Understanding, vol. 210, p. 103241, 2021, doi: 10.1016/j.cviu.2021.103241.

[6] C. Singh, E. Walia, and K. P. Kaur, “Color texture description with novel local binary patterns for effective image retrieval,” Pattern Recognition, vol. 76, pp. 50–68, 2018, doi: 10.1016/j.patcog.2017.10.021.

[7] I. Al Saidi, M. Rziza, and J. Debayle, “A Novel Texture Descriptor: Circular Parts Local Binary Pattern,” Image Analysis and Stereology, vol. 40, no. 2, pp. 105–114, 2021, doi: 10.5566/ias.2580.

[8] S. Fadaei, P. Hosseini, and K. RahimiZadeh, “New Texture Descriptor Based on Improved Orthogonal Difference Local Binary Pattern,” in Proc. 6th Int. Conf. Pattern Recognition and Image Analysis (IPRIA), 2023, pp. 1–5, doi: 10.1109/IPRIA59240.2023.10147180.

[9] A. Arivarasan and K. M, “Determination of Color and Shape Features Based Image Retrieval with Hybrid Feature Descriptor Construction (HFDC),” Indian Journal of Science and Technology, vol. 12, no. 42, pp. 1–12, 2019, doi: 10.17485/ijst/2019/v12i42/148143.

[10] S. Ramasinghe, S. Khan, and N. Barnes, “Learned and Hand-crafted Feature Fusion in Unit Ball for 3D Object Classification,” in Proc. 9th Int. Conf. Pattern Recognition Applications and Methods, 2020, pp. 115–125, doi: 10.5220/0009344801150125.

[11] Sutikno, H. A. Wibawa, and I. Waspada, “Feature Fusion of GLCM, LBP, and Color Histogram for Rice Leaf Disease Classification,” Ingénierie des Systèmes d’Information, vol. 31, no. 1, 2026, doi: 10.18280/isi.310101.

[12] R. Kumudham and V. Rajendran, “Classification performance assessment in side scan sonar image while underwater target object recognition using random forest classifier and support vector machine,” International Journal of Engineering & Technology, vol. 7, no. 2.21, p. 386, 2018, doi: 10.14419/ijet.v7i2.21.12448.

[13] M. F. Kurniawan and D. A. Megawaty, “Comparison of Logistic Regression, Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) Algorithms in Diabetes Prediction,” Journal of Applied Informatics and Computing, vol. 9, no. 5, pp. 2154–2162, 2025, doi: 10.30871/jaic.v9i5.9815.

[14] S. Sudharson, R. Annamalai, A. A. Reddy, and P. Varsha, “Stacking Ensemble Model for Celestial Object Classification: Galaxies, Stars and Quasars,” in Proc. Seventh Int. Conf. Image Information Processing (ICIIP), 2023, pp. 746–752, doi: 10.1109/ICIIP61524.2023.10537787.

[15] G. Yildiz, A. Ulu, B. Dızdaroglu, and D. Yildiz, “Hybrid Image Improving and CNN (HIICNN) Stacking Ensemble Method for Traffic Sign Recognition,” IEEE Access, vol. 11, pp. 69536–69552, 2023, doi: 10.1109/ACCESS.2023.3292955.

[16] G. Wang, Y. Guo, Z. Xu, and M. Kankanhalli, “Bilateral Adaptation for Human-Object Interaction Detection with Occlusion-Robustness,” in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition, 2024, pp. 27970–27980, doi: 10.1109/CVPR52733.2024.02642.

[17] W. Bejjani, W. C. Agboh, M. R. Dogar, and M. Leonetti, “Occlusion-Aware Search for Object Retrieval in Clutter,” in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems, 2021, pp. 4678–4685, doi: 10.1109/IROS51168.2021.9636230.

[18] E. U. Samani and A. G. Banerjee, “Persistent Homology Meets Object Unity: Object Recognition in Clutter,” IEEE Transactions on Robotics, vol. 40, pp. 886–902, 2024, doi: 10.1109/TRO.2023.3343994.

[19] P. L, P. K, G. V, and K. P, "Improved Occlusion Handling in Object Recognition by Enhancing Image Quality and Suitable Illumination Techniques," in Proc. 5th Int. Conf. Mobile Computing and Sustainable Informatics (ICMCSI), 2024, pp. 218–224, doi: 10.1109/ICMCSI61536.2024.00039.

[20] A. Chand and R. S. Madhan, "A Robust Shape-Based Hybrid Feature Stacking Ensemble for Object Recognition on COIL-100," International Journal of Applied Mathematics, vol. 38, no. 12s, 2025.