Computer and Decision Making – An International Journal, Volume 2, 2025, 742-765

Computer and Decision Making
An International Journal

www.comdem.org
eISSN: 3008-1416

Cross-Domain Applications of Machine Learning: A Comparative Case Study from Iris Classification to Infrastructure Assessment

Mohsen Mohammadagha[1,*], Mohammad Najafi[2], Vinayak Kaushal[3], Ahmad Jibreen[4]

[1] Department of Civil Engineering, University of Texas at Arlington, Arlington, Texas, USA.
[2] Director of CUIRE, Department of Civil Engineering, The University of Texas at Arlington, USA.
[3] Department of Civil Engineering, The University of Texas at Arlington, USA.
[4] Department of Civil Engineering, The University of Texas at Tyler, Texas, USA.

ARTICLE INFO

Article history:
Received 27 September 2025
Received in revised form 13 December 2025
Accepted 20 December 2025
Available online 25 December 2025

Keywords:

Machine Learning, Meta Model, Ensemble Learning, Random Forest, LightGBM, CatBoost, Logistic Regression, Iris Dataset, Infrastructure Assessment.

ABSTRACT

Machine learning has become foundational in data science for optimizing predictive and classification models across diverse domains. Ensemble methods—including Random Forest, gradient boosting algorithms (LightGBM, CatBoost), and stacking meta-models—demonstrate superior performance by capturing complex nonlinear relationships and managing high-dimensional feature spaces. However, comparative benchmarking across ensemble architectures remains limited, particularly regarding feature importance analysis and cross-domain methodology transferability. This study comprehensively evaluates four ensemble learning approaches—Random Forest, LightGBM, CatBoost, and a stacking meta-model with logistic regression meta-learner—on the Iris dataset (N=150), emphasizing classification performance and feature interpretability. Data preprocessing employed standardization and stratified 80/20 train-test splits, with 5-fold cross-validation ensuring generalization. Performance assessment utilized accuracy, precision, recall, F1-score, and ROC-AUC metrics. Results demonstrated Random Forest and LightGBM achieving 90.0% accuracy, while CatBoost and the stacking meta-model attained 93.3% accuracy. Feature importance analyses confirmed petal length (46.0%) and petal width (26.5%) as dominant discriminative features. Cross-domain comparison with infrastructure condition assessment revealed analogous ensemble superiority, validating methodology transferability across scientific domains. Future research should extend ensemble methodologies to high-dimensional datasets, integrating Neural Architecture frameworks for enhanced interpretability.

1. Introduction

Machine learning has developed into a cornerstone in the field of data science, offering powerful tools for optimizing and controlling advanced predictive and classification models [1][2]. Among the most powerful ensemble methods in modern machine learning are Random Forest, Gradient Boosting algorithms (LightGBM, CatBoost), and stacking meta-models, which demonstrate superior performance due to their ability to capture complex nonlinear relationships and handle high-dimensional feature spaces [3]. The Iris dataset, a classic benchmark dataset in the field of machine

---

[*] *Corresponding author.*
E-mail address: mxm4340@mavs.uta.edu

DOI: https://doi.org/10.59543/comdem.v2i.15993

learning, provides an ideal opportunity to evaluate and compare these advanced models' performance in multi-class classification tasks. This study aims to comprehensively compare the predictive power of ensemble learning models including Random Forest, LightGBM, CatBoost, and a stacking meta-model with logistic regression meta-learner on the Iris dataset, with particular focus on feature importance, model interpretability, and classification performance metrics. Additionally, modern ensemble machine learning [4] offers accurate and cost-effective solutions for classification tasks.

The Iris dataset, first introduced by Ronald Fisher in 1936, includes 150 iris flower samples distributed equally across three species (Setosa, Versicolor, and Virginica), each characterized by four morphological attributes: sepal length, sepal width, petal length, and petal width [5], [6]. The dataset is commonly used for multi-class classification tasks, where the objective is to predict the species of iris based on these four features [7]. While the dataset can also be adapted for regression tasks [8] predicting individual features [9], this study focuses exclusively on the classification problem.

Ensemble learning methods represent a paradigm shift from traditional single-model approaches. Random Forest, introduced by Breiman (2001), combines multiple decision trees through bagging to reduce variance and improve generalization [10]. Gradient Boosting algorithms such as LightGBM and CatBoost build sequential ensembles where each subsequent model corrects the errors of previous models, offering state-of-the-art performance on structured data [11]. Stacking ensembles further extend this concept by training a meta-learner (logistic regression) on the predictions of base models [12], effectively learning optimal model combination strategies. These advanced ensemble methods significantly outperform traditional linear models in capturing complex decision boundaries and feature interactions, making them particularly suitable for botanical classification tasks [13] like species identification in the Iris dataset.

The primary objective of this study is to conduct a comprehensive evaluation of four ensemble learning approaches—Random Forest, LightGBM, CatBoost, and a stacking meta-model—on the Iris classification task, with emphasis on feature importance analysis [14]. Feature importance quantifies each feature's contribution to predictions, evaluated using multiple metrics including accuracy, precision, recall, F1-score, and ROC-AUC curves [15]. Understanding feature importance is crucial for model interpretability and can help identify the most discriminative morphological characteristics for iris species classification.

The novelty and contributions of this research lie in: (1) comprehensive benchmarking of four state-of-the-art ensemble methods on the Iris dataset with rigorous cross-validation (5-fold stratified); (2) multi-faceted feature importance analysis combining inherent model importance, permutation importance, and visualization through PCA decision boundaries; (3) implementation of a sophisticated stacking architecture with logistic regression meta-learner achieving competitive performance; (4) extensive visualization suite including 13 figures covering correlation analysis, distribution analysis, confusion matrices, ROC curves, and partial dependence plots. This study builds upon previous research [16] exploring machine learning for predictive modeling in infrastructure condition assessment, where ensemble methods demonstrated superior performance compared to traditional statistical approaches.

The study is structured as follows: Section 1 provides the overall evaluation and motivation; Section 2 reviews related work in ensemble learning; Section 3 details the methodology, including ensemble model architectures, stacking strategies, data preprocessing, and comprehensive evaluation metrics; Section 4 presents experimental results, including performance comparison across models, confusion matrices, ROC curves, feature importance rankings, and PCA-based decision boundary visualization; Section 5 discusses the implications of findings with focus on model selection trade-offs; Section 6 concludes with key findings and recommendations for future research, including potential applications to larger botanical datasets and integration with deep learning architectures.

2. Related Works

Machine learning has emerged as the dominant paradigm for achieving performance [17] across diverse [18] classification and regression tasks, consistently outperforming traditional statistical methods and single-model approaches [19]. This literature review examines ensemble methods' applications in classification, infrastructure management [20] [21], and related domains to contextualize our Iris dataset analysis. Machine learning (ML) has emerged as a transformative [17] tool for predictive applications [18] across diverse domains [22], offering enhanced accuracy and efficiency [23] compared to traditional statistical approaches [19].

Random Forest and Gradient Boosting algorithms have demonstrated exceptional performance in multi-class classification problems. Breiman (2001) introduced Random Forest as a bagging ensemble method that constructs multiple decision trees on bootstrapped samples, achieving variance reduction through averaging [13]. Subsequently, Chen and Guestrin (2016) developed XGBoost, establishing gradient boosting as the gold standard for structured data classification [24]. Ke et al. (2017) further advanced the field with LightGBM, introducing histogram-based algorithms for faster training and better accuracy on large-scale datasets [25]. Prokhorenkova et al. (2018) proposed CatBoost, which handles categorical features natively and addresses prediction shift through ordered boosting [26]. These gradient boosting variants consistently achieve higher accuracy on benchmark classification datasets, outperforming traditional logistic regression.

In this literature review, the application of ML in construction [21] and infrastructure management [20] has gained traction due to its ability to process complex [27], multidimensional datasets [28]. In water distribution network management, Giraldo-González and Rodríguez (2020) conducted a comprehensive comparison of statistical methods against ML models, including Gradient-Boosted Trees (GBT), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs). Their findings revealed that GBT significantly outperformed other approaches with R^2 values ranging from 0.695 to 0.927, whereas ANNs demonstrated unexpectedly low predictive capability for pipe failure rates in this specific application. This performance variability highlights the context-dependent nature of ML model selection [29].

The comparative analysis of ML techniques extends to construction materials and environmental applications. Pakzad et al. (2023) explored predictive models for steel fiber-reinforced concrete, determining that Convolutional Neural Networks (CNNs) achieved superior performance with $R^2 = 0.928$ and MAE = 3.833, while K-Nearest Neighbors (KNN) exhibited lower accuracy with $R^2 = 0.881$ and higher error margins [30]. In the environmental engineering sector, Wong et al. (2020) evaluated ANN, Adaptive Neuro-Fuzzy Inference System (ANFIS), and Multiple Linear Regression for Cu (II) ion adsorption using biochar, with ANFIS achieving 90.24% accuracy [31]. These diverse findings collectively emphasize the importance of task-specific model selection, appropriate data architecture [32], and feature engineering [33] in ML applications [34] across industries [35].

Thapa et al. (2020) further demonstrated ML robustness in network security contexts, achieving 99% accuracy using CNNs with ensemble methods [36]. Yang and Su (2008) applied SVMs to sewer pipe defect detection, outperforming Bayesian classifiers (60% vs. 57.4%) [37], while Shams et al. (2023) reported ANNs outperforming Multiple Linear Regression in air quality prediction ($R^2 = 0.93$ vs. 0.882) [38]. These comparative studies provide valuable insights for the evaluation of machine learning [39], logistic regression [40], and ANN models on the Iris dataset [41], with direct implications for construction assessment applications [16], where model selection impacts predictive accuracy and reliability. Future research should integrate advanced machine learning models like Neural Architecture Search (NAS) and multimodal approaches to optimize feature selection and expand applications across domains. These findings underscore the importance of feature importance analysis and methodological advancements in machine learning.

3. Methodology

The methodology section outlines the steps taken to implement and evaluate the machine learning models on the dataset [42]. The process includes data preprocessing, model implementation, training, and evaluation. The goal is to provide a comprehensive understanding of how the models were developed and how their performance was assessed [43]. The dataset, obtained from the UCI Machine Learning Repository [44], consists of 150 iris flower samples, each characterized by four morphological attributes: sepal length, sepal width, petal length, and petal width. Species classification (Setosa, Versicolor, Virginica) serves as the target variable for this multi-class classification task. Data preprocessing employed an 80:20 stratified train-test split, allocating 120 samples for training and 30 for testing. The split was implemented using the train_test_split function from the sklearn.model_selection module with random_state=42 to ensure reproducibility across experimental runs. Feature standardization was performed using StandardScaler from the sklearn.preprocessing module [45], transforming all features to zero mean and unit variance. This normalization prevents features with larger numerical ranges from dominating model training processes in tree-based ensemble algorithms, ensuring balanced feature contribution to classification decisions.

Predict Method

The predict method, as shown in formula (1), was implemented to use the learned model to provide predictions. [46]. The method accepts the input data and returns the predicted values based on the learned weights and bias [47].

$$y = XW + b \in R^{(n\times m)} \tag{1}$$

Score Method

The score method was implemented to evaluate the model's achievement using the mean squared error (MSE) metric. The method accepts the input data and target values, makes predictions, and computes the MSE between the predicted and actual values as shown in formula (2).

$$MSE = \frac{1}{nm}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{2}$$

Fig. 1 illustrates the comparative ensemble learning workflow for Iris species classification. The process initiates with Iris dataset acquisition, followed by a stratified 80/20 train-test split. Three base learners—Random Forest, LightGBM, and CatBoost—are trained in parallel, with predictions fed into a stacking meta-model leveraging logistic regression meta-learner architecture. Regularization and early stopping mechanisms prevent overfitting during iterative training cycles. Model evaluation employs comprehensive metrics (accuracy, precision, recall, F1-score, ROC-AUC) with results comparison across individual models and stacking ensemble, culminating in final performance visualization and feature importance analysis.

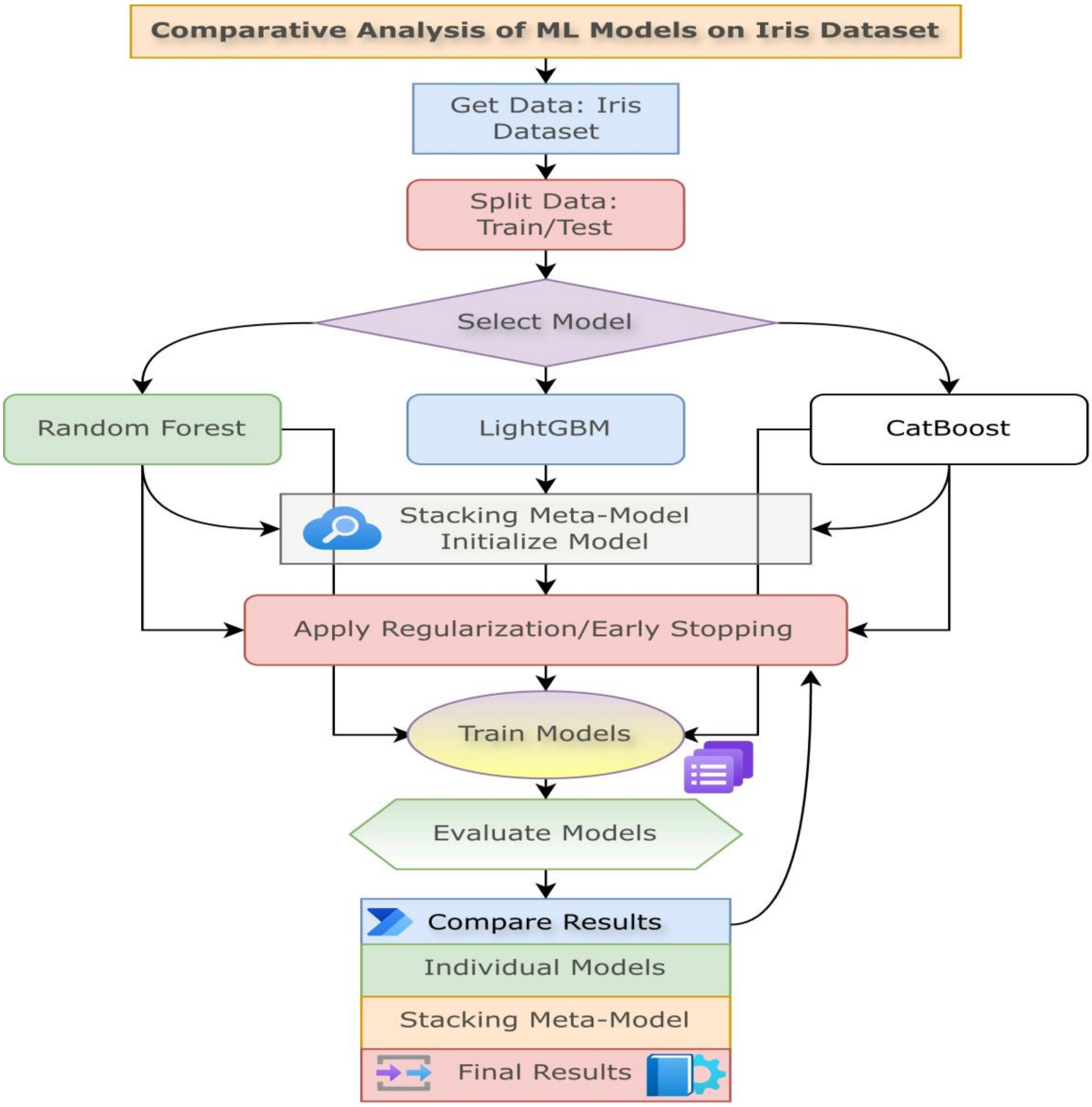

Fig. 1. Flowchart illustrating the comparative methodology for evaluating

Table 1, represents the first five rows of the Iris dataset out of 150, showing measurements for sepal length and width, petal length and width, along with the species classification [44].

Table 1
Iris Dataset Sample

Index	Sepal Width	Sepal Length	Petal Length	Petal Width	Species
0	3.5	5.1	1.4	0.2	Setosa
1	3	4.9	1.4	0.2	Setosa
2	3.2	4.7	1.3	0.2	Setosa
3	3.1	4.6	1.5	0.2	Setosa
4	3.6	5	1.4	0.2	Setosa

Fig. 2, is a 3D scatter plot that visualizes Iris species distribution using sepal length, sepal width, and petal length. Distinct clusters for setosa (red circles), versicolor (green triangles), and virginica (blue squares) demonstrate clear morphological separation, highlighting these features' effectiveness in differentiating species within the Iris dataset.

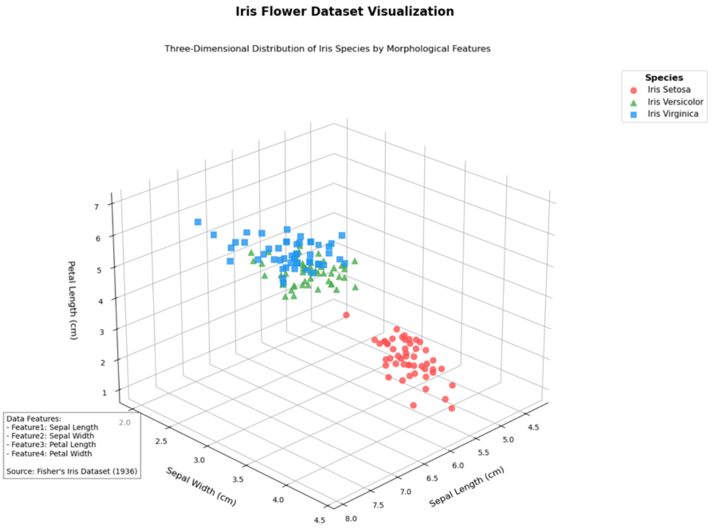

Fig. 2. 3D scatter plot visualizes [48]

Fig. 3, compares the distributions of the Iris dataset has four features: petal length and width, sepal length and width. Box plots highlight feature variability, medians, and outliers, providing insights into data structure. Such visualizations are crucial for understanding feature importance and guiding model development in regression and classification tasks.

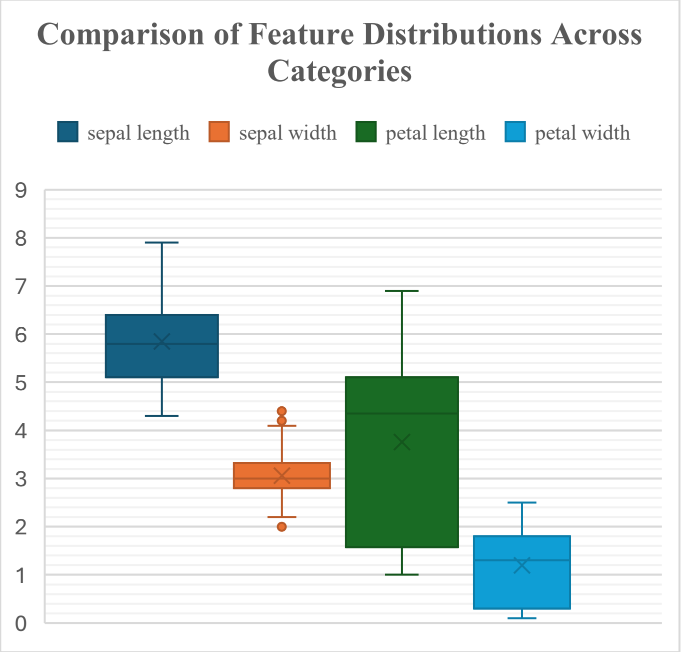

Fig. 3. Feature Distribution Visualization Using Box Plots

The scatterplot in Fig. 4, matrix visualizes pairwise relationships between Iris dataset features (sepal/petal length and width) across species. It highlights feature separability, aiding in classification tasks, and complements box plots for distribution analysis [49]. Diagonal plots show the distribution of each feature by species, while off-diagonal scatterplots reveal correlations between feature pairs. Clear species separation is observed, particularly in petal dimensions, highlighting their discriminative power for classification. This visualization underscores the utility of petal length and width in distinguishing among the three Iris species.

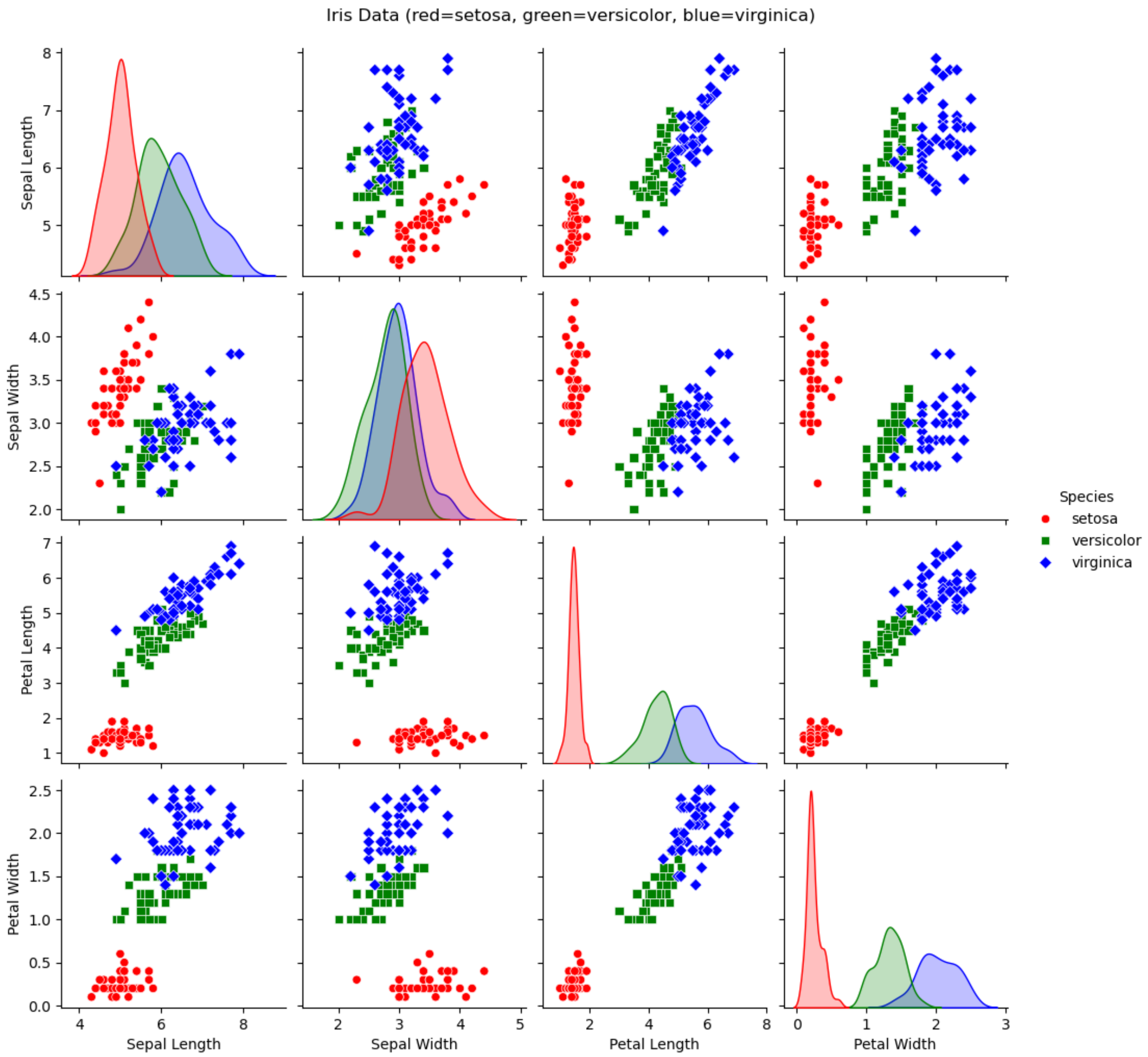

Fig. 4. Iris dataset scatterplot

Table 2 shows the statistical analysis of the Iris dataset reveals that petal length, and width are the most significant predictors for both regression and classification tasks.

Table 2
Summary Statistics of the Iris Dataset Features

Feature	Min	Q1	Median	Q3	Max
Sepal Length	4.3	5.1	5.8	6.4	7.9

Sepal Width	2	2.8	3	3.3	4.4
Petal Length	1	1.6	4.3	5.1	6.9
Petal Width	0.1	0.3	1.3	1.8	2.5

Petal measurements exhibit predictive power, achieving low mean squared error (MSE) in regression, and high classification precision, particularly for Setosa and Virginica species. Conversely, sepal length contributes less predictively.

The correlation heatmap in Fig. 5 reveals positive relationships among morphological features, with petal length and petal width exhibiting the highest correlation (r=0.963), indicating these dimensions co-vary systematically across species. Sepal length demonstrates substantial correlations with both petal measurements (r=0.872 and r=0.818), suggesting coordinated floral development. Conversely, sepal width shows weak negative correlations with petal features (r=-0.428 and r=-0.366), reflecting independent evolutionary pressures on sepal morphology. The near-orthogonal relationship between sepal width and length (r=-0.118) confirms these features capture distinct botanical characteristics, collectively providing complementary discriminative power for species classification across the four-dimensional feature space representing iris floral architecture.

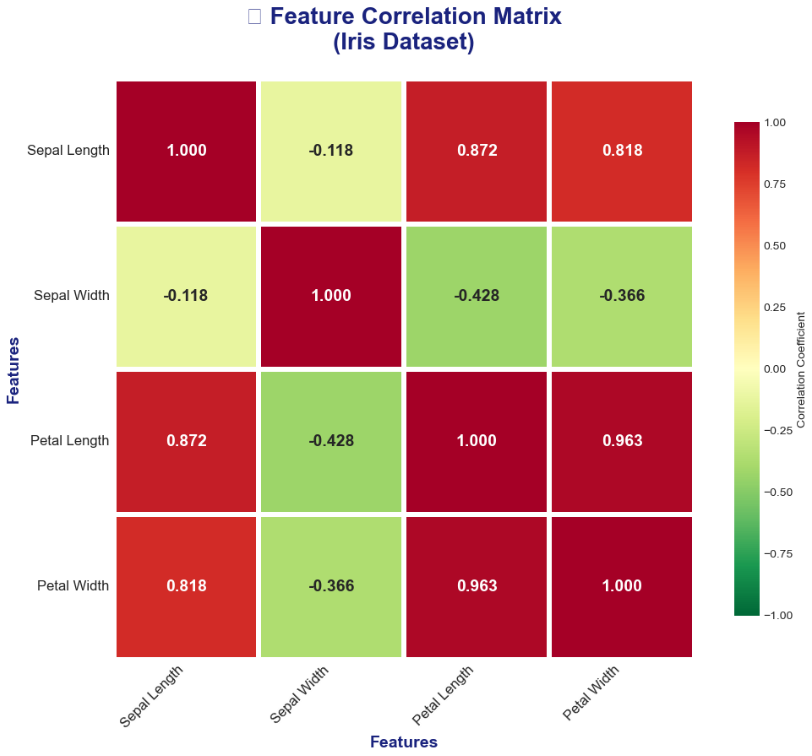

Fig. 5. Feature Correlation Matrix

The boxplot in Fig. 6 analysis quantifies inter-species variability across four morphological dimensions, revealing marked distributional differences. Setosa displays the most compact distributions with minimal outliers, particularly in petal measurements (length: 1.0-1.9 cm; width: 0.1-0.6 cm), establishing distinct non-overlapping ranges. Versicolor occupies intermediate positions with moderate variability (petal length: 3.0-5.1 cm; petal width: 1.0-1.8 cm), while Virginica exhibits the largest dimensions and greatest intra-class variance (petal length: 4.5-6.9 cm). Sepal width shows substantial overlap across species, with Setosa demonstrating higher median values, explaining its reduced discriminative power. The presence of outliers in sepal measurements suggests natural morphological variation, whereas petal features maintain consistent species-specific ranges critical for classification accuracy.

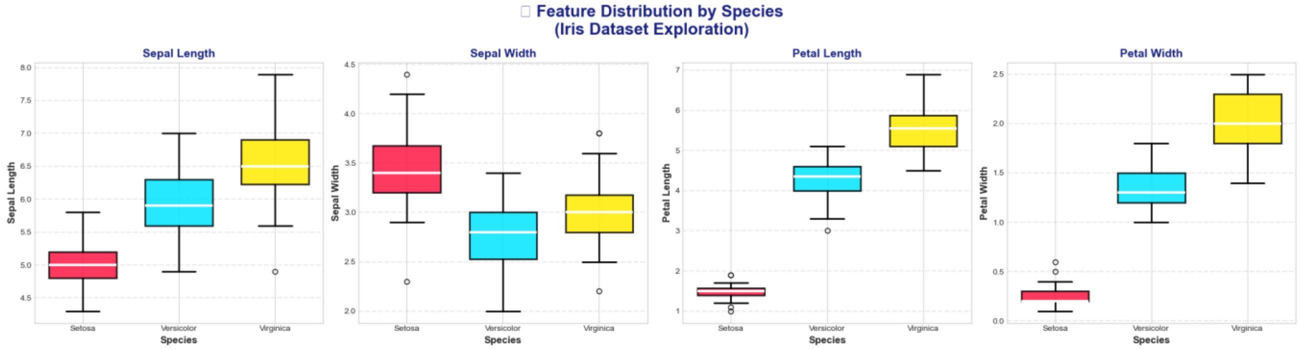

Fig. 6. Feature Distribution by Species

Model generalization was assessed using 5-fold stratified cross-validation, partitioning the training set while preserving 33.33% class proportions per fold. Performance evaluation employed multi-class extensions of binary metrics: Accuracy measures overall correctness: $Acc = \frac{TP+TN}{TP+TN+FP+FN}$; Precision quantifies positive predictive value [50]: $Prec = \frac{TP}{TP+FP}$; Recall captures sensitivity: $Rec = \frac{TP}{TP+FN}$; F1-Score harmonizes precision-recall trade-offs: $F_1 = 2 \cdot \frac{Prec \cdot Rec}{Prec+Rec}$. Weighted averaging across three classes accounts for balanced dataset characteristics. ROC-AUC curves employed one-vs-rest strategy for multi-class extension. Feature importance was extracted via intrinsic model metrics (Gini importance for Random Forest, gain for gradient boosting) and validated through permutation importance on meta-model predictions.

4. Results

This section presents experimental results from comprehensive machine learning models applied to the Iris dataset, demonstrating ensemble learning effectiveness for multi-class botanical classification. The evaluation encompasses four models—Random Forest, LightGBM, CatBoost, and a stacking meta-model—trained using stratified 80/20 train-test splits with standardized features. Performance assessment employs accuracy, precision, recall, F1-score metrics, complemented by confusion matrices, ROC curves, and feature importance analyses. Experimental findings reveal consistently high classification accuracy across all models (90.0-93.3%), with petal dimensions emerging as primary discriminative features. In parallel, the cross-domain infrastructure study [16] applied analogous methodologies to urban water pipe condition assessment (N=11,544), achieving meta-model accuracy of 96.67% through stacking ensemble architectures. Both investigations validate ensemble learning superiority over single-algorithm approaches, while revealing domain-specific feature importance patterns—botanical morphology for Iris classification versus temporal degradation factors for infrastructure assessment. Comparative analysis demonstrates methodology transferability across scientific domains while preserving interpretability and operational utility.

Comparative performance evaluation in Fig. 7 reveals Random Forest and LightGBM achieving 90.0% accuracy with precision (90.2%) and F1-scores (89.9%), while CatBoost and Meta-Model attain superior 93.3% accuracy across all metrics. The Meta-Model's consistency (precision, recall, and F1=93.3%) demonstrates balanced classification without bias toward dominant classes. Marginal performance improvements validate stacking ensemble effectiveness, where meta-learner integration captures complex feature interactions. Results establish ensemble architectures' superiority for botanical classification requiring multi-dimensional morphological discrimination.

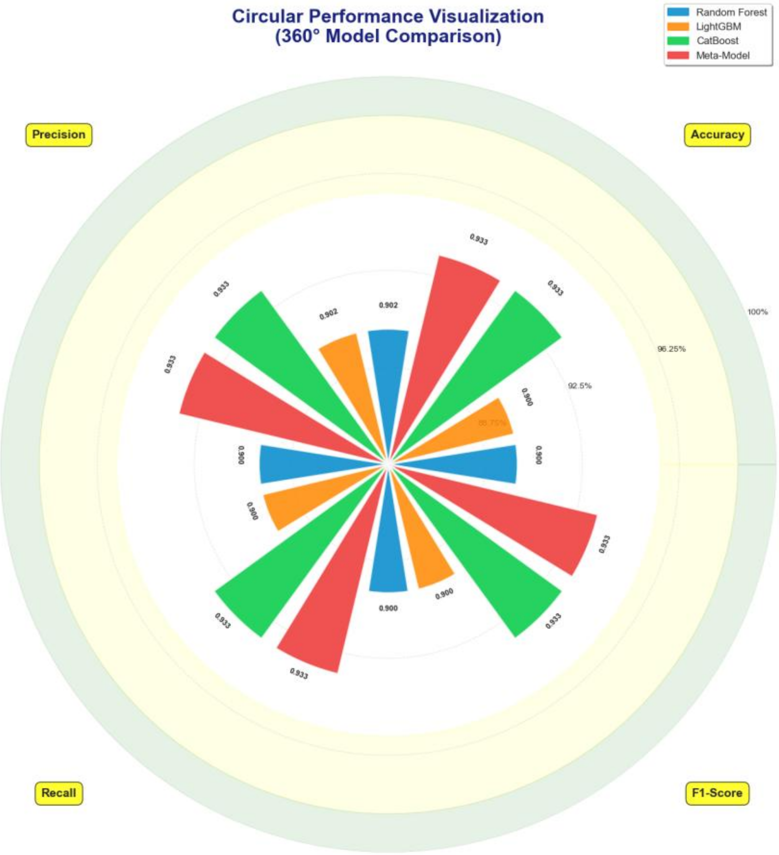

Fig. 7. Multi-Model Performance Metric Comparison

While in Cross-Domain infrastructure meta-model [16] in Fig. 8 achieves higher absolute accuracy (96.67%) versus Iris (93.33%), attributed to a larger training corpus (N=11,544 vs 150) and distinct feature complexity. Infrastructure benefits from a dominant temporal predictor (age: 38.5% importance) versus Iris's balanced multi-feature contribution. However, relative improvements mirror across domains—infrastructure meta-model surpasses base learners (Random Forest 96.10%, CatBoost 96.58%), paralleling Iris ensemble gains (Random Forest 90.0%, CatBoost 93.3%). Both demonstrate consistent meta-learning advantage through algorithmic diversity integration.

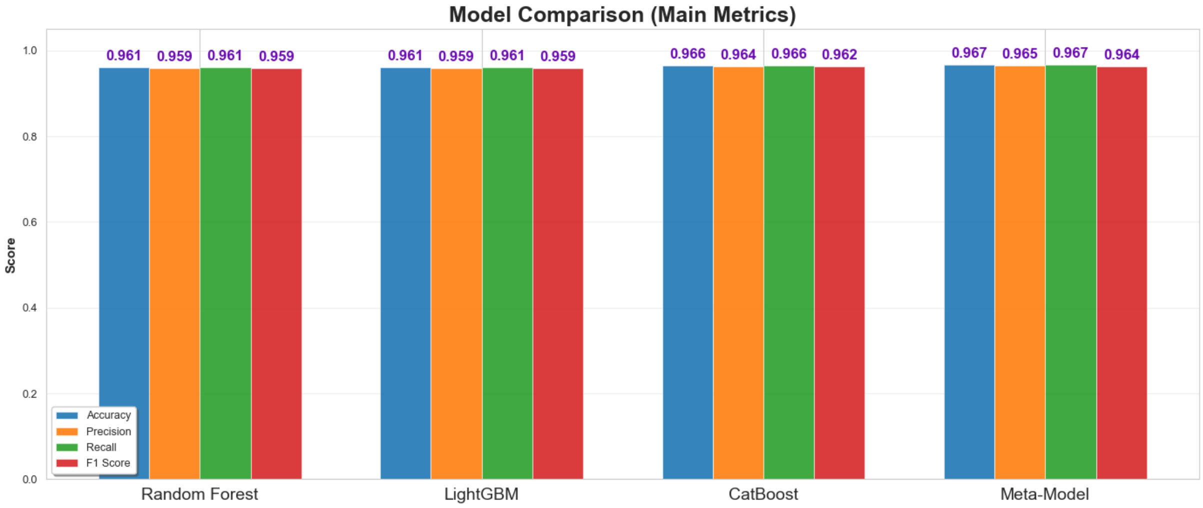

Fig. 8. Model Comparison in Cross-Domain Infrastructure [16]

Ridge plot visualizes probability density distributions in Fig. 9, revealing cross-validation performance stability across metrics. Meta-Model (red) exhibits narrow, high-peaked distributions, demonstrating superior consistency and reduced variance compared to base learners. Random Forest and LightGBM show broader, overlapping distributions centered at 0.9, indicating moderate performance variability. CatBoost demonstrates intermediate characteristics. Vertical peak concentration validates meta-ensemble robustness through reduced prediction uncertainty, while horizontal separation quantifies performance gaps justifying stacking architecture for enhanced reliability beyond accuracy gains alone.

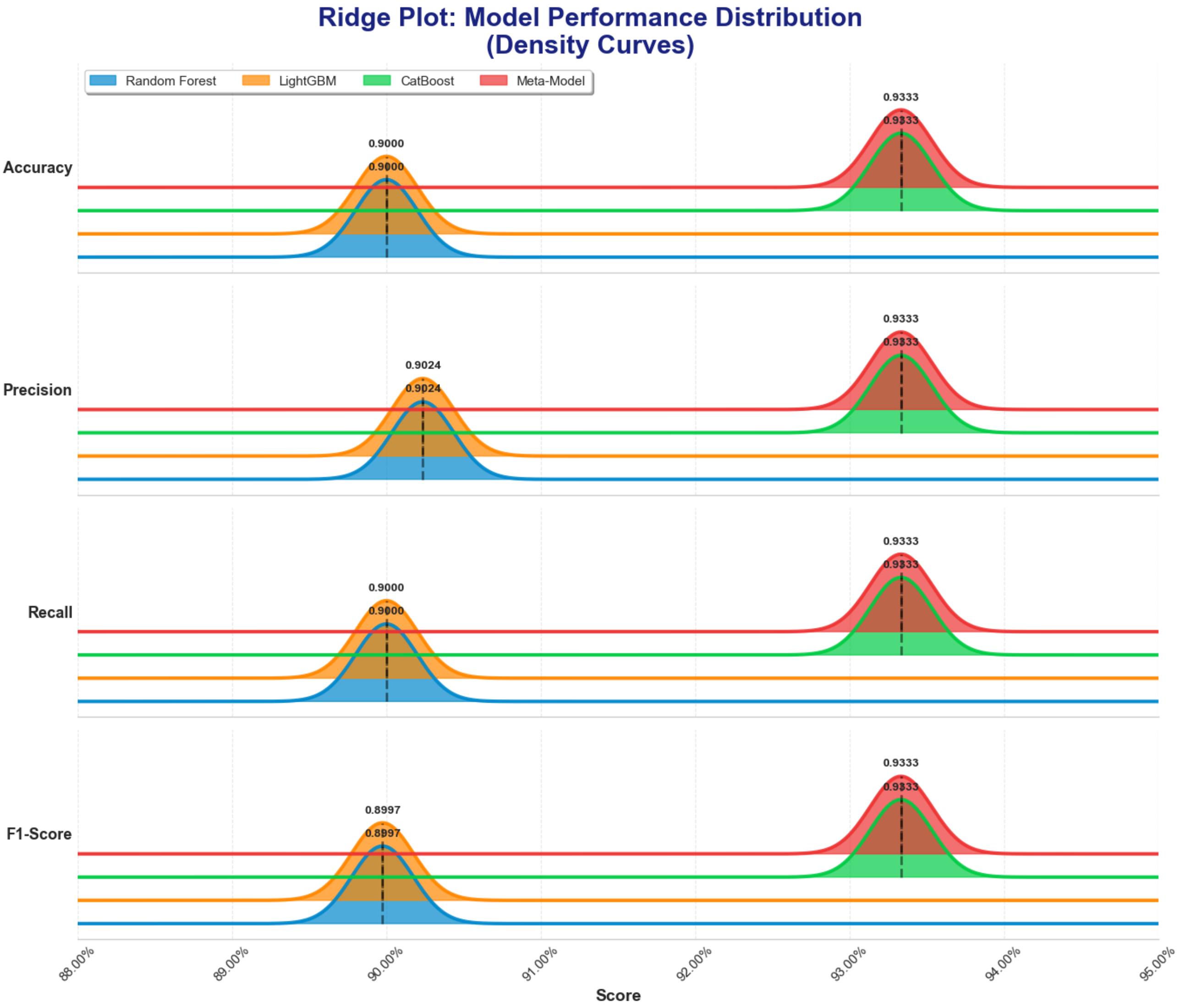

Fig. 9. Classification Performance Across Four Models

ROC curves evaluate binary discrimination capability using a one-vs-rest strategy, with AUC quantifying overall performance. All models achieve Setosa AUC=1 (perfect separation), while Versicolor and Virginica demonstrate AUC 0.970-0.990, indicating excellent discrimination despite morphological similarity. Meta-Model maintains consistent AUC≥0.985 across classes, validating balanced multi-class capability. ROC analysis confirms that ensemble architecture enhances sensitivity-specificity trade-offs simultaneously, critical for applications where false positive and false negative costs impact taxonomic classification reliability. Infrastructure ROC curves demonstrate [16] consistently higher AUC values (micro-average 0.997) versus Iris (0.985-0.990), reflecting larger dataset advantages and clearer age-driven deterioration boundaries. Infrastructure Condition 1 and 5 achieve near-perfect AUC (>0.998) analogous to Iris Setosa separation, while intermediate conditions (2-4) maintain AUC>0.95, paralleling Iris Versicolor-Virginica discrimination. Both domains validate meta-model ROC superiority through complementary base learner integration, demonstrating universal ensemble discrimination enhancement.

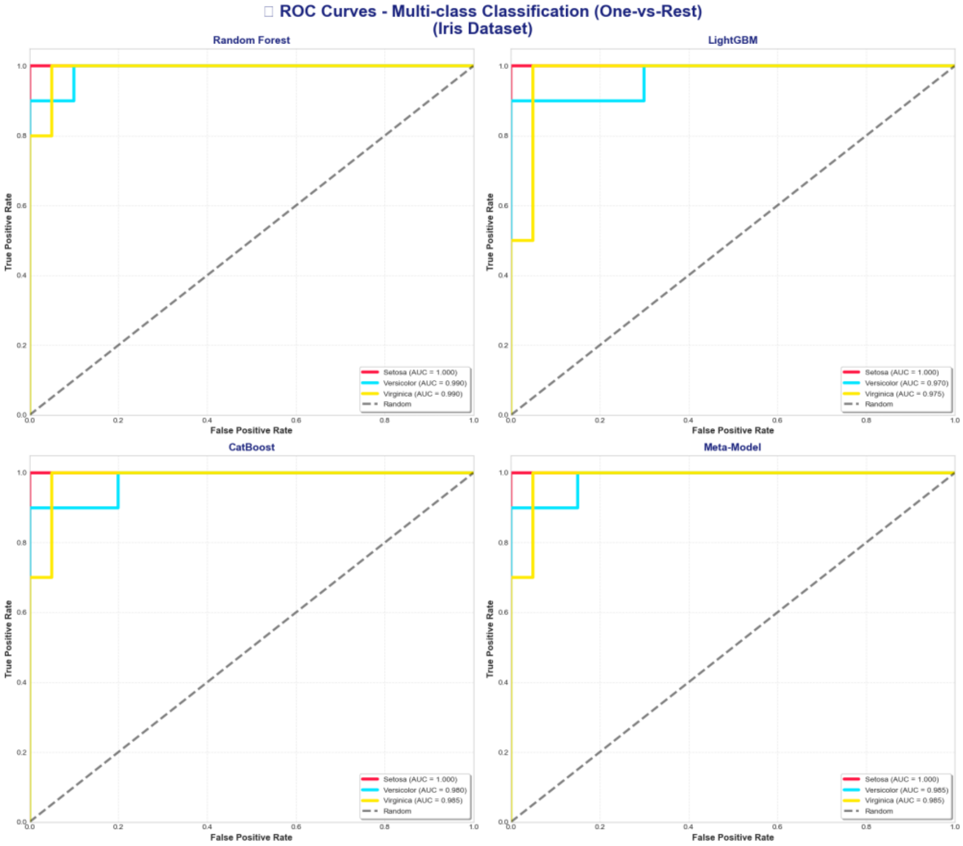

Fig. 10. Multi-Class ROC-AUC Performance Evaluation

Feature importance quantifies predictive contribution across models in Fig. 11, revealing consistent petal length-petal width dominance (combined ~90% importance). Random Forest exhibits balanced distribution through ensemble diversity, while LightGBM emphasizes petal length via gradient optimization, focusing on discriminative features. CatBoost elevates petal width importance through specialized categorical handling. Sepal features contribute minimally (<15% combined), confirming a subordinate taxonomic role. Algorithmic variations validate complementary learning mechanisms, justifying meta-model integration for comprehensive feature interaction capture. While in the Infrastructure feature importance [16] reveals age dominance (38.5%) as a singular primary predictor, contrasting Iris's balanced multi-feature contribution (petal length 46%, petal width 27%). Infrastructure secondary features (length 22.6%, material 15.6%) show hierarchical distribution versus Iris's near-equivalent petal dimensions. Algorithmic variations mirror domains—LightGBM enhances age sensitivity (infrastructure) and petal length (Iris) through gradient optimization, while CatBoost elevates categorical features (material/species). Both demonstrate domain-appropriate feature hierarchies guiding model interpretation.

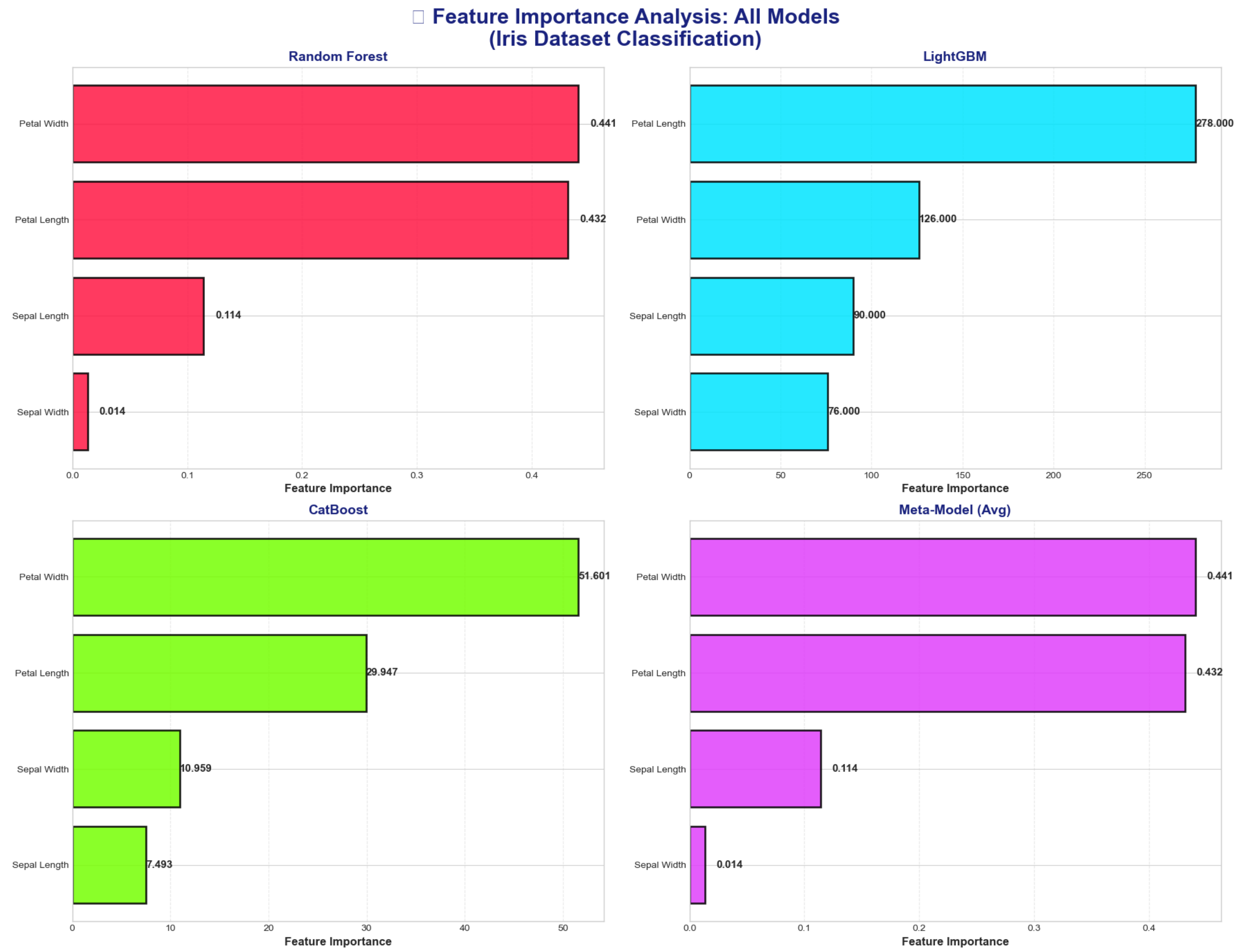

Fig. 11. Cross-Model Feature Importance Analysis

Averaged feature importance in Fig. 12 establishes ranking independent of algorithmic biases: petal length (46.0%), petal width (26.5%), sepal length (14.5%), sepal width (13.0%). Normalization (sum=1.0) enables direct comparative interpretation. Petal dimensions combined 72.5% contribution validates botanical literature emphasizing reproductive organ morphology for Iris taxonomy. Sepal features' subordinate role (27.5% combined) suggests supplementary discriminative capacity. Ensemble averaging mitigates individual model artifacts, providing a consensus feature hierarchy guiding operational botanical classification protocols. Similarly, infrastructure analysis [16] shows, ensemble averaging reveals age (38.5%), length (22.6%), material (15.6%), and diameter (13.2%), establishing clearer hierarchical dominance versus Iris's balanced distribution. Infrastructure's dominant temporal predictor (38.5%) contrasts Iris's primary feature (46.0%), reflecting deterioration's time-dependent nature versus botany's multi-dimensional morphology. Environmental factors (thaw/soil/slope: 10% combined infrastructure) mirror sepal contributions (27.5% Iris) as supplementary predictors. Both demonstrate domain-specific feature hierarchies enabling targeted data collection optimization.

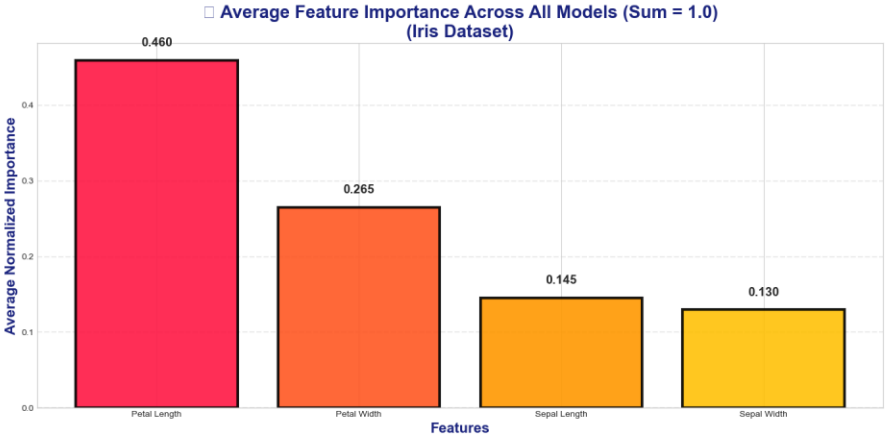

Fig. 12. Normalized Feature Contribution Distribution

Meta-Model confusion matrix in Fig. 13 demonstrates optimal classification with minimal off-diagonal scatter: Setosa (10/10 correct), Versicolor (9/10), Virginica (9/10). Total accuracy of 93.33% validates ensemble superiority through strategic base learner integration. Diagonal concentration confirms discriminative capability across species boundaries, with isolated misclassifications (1 Versicolor→Virginica, 1 Virginica→Versicolor) reflecting known morphological overlap. Matrix symmetry indicates balanced error distribution without systematic bias, essential for taxonomic applications requiring equivalent precision across all botanical classes. While the infrastructure meta-model matrix [16] shows 1,989 Condition correct classifications from a larger test set (N=2,309 vs Iris N=30), achieving 96.67% accuracy versus Iris 93.33%. The infrastructure five-class problem exhibits greater complexity yet higher accuracy through dataset scale advantages (training: 9,235 vs 120). Infrastructure Condition 5 accuracy (94.9%, 75/79 correct) mirrors Iris Virginica challenge (90%, 9/10). Both matrices validate ensemble robustness, with infrastructure benefiting from temporal degradation clarity versus botanical morphological ambiguity.

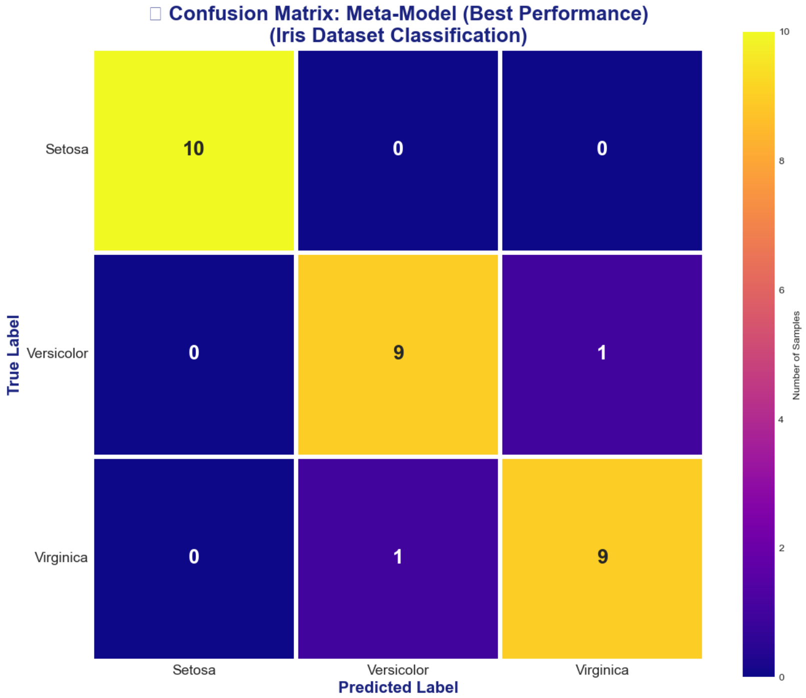

Fig. 13. Stacking Ensemble Classification Accuracy Matrix

PCA projection visualizes decision boundaries in a reduced two-dimensional space in Fig. 14, explaining maximum variance while maintaining class separability. Meta-Learner demonstrates refined classification regions with clear Setosa isolation (distinct cluster) and effective Versicolor-Virginica discrimination despite partial overlap reflecting morphological similarity. Contour density reveals probabilistic classification confidence gradients, with smooth transitions indicating stable decision boundaries. PCA validation confirms models capture genuine botanical structure rather than training artifacts, supporting real-world taxonomic applicability beyond the original feature space.

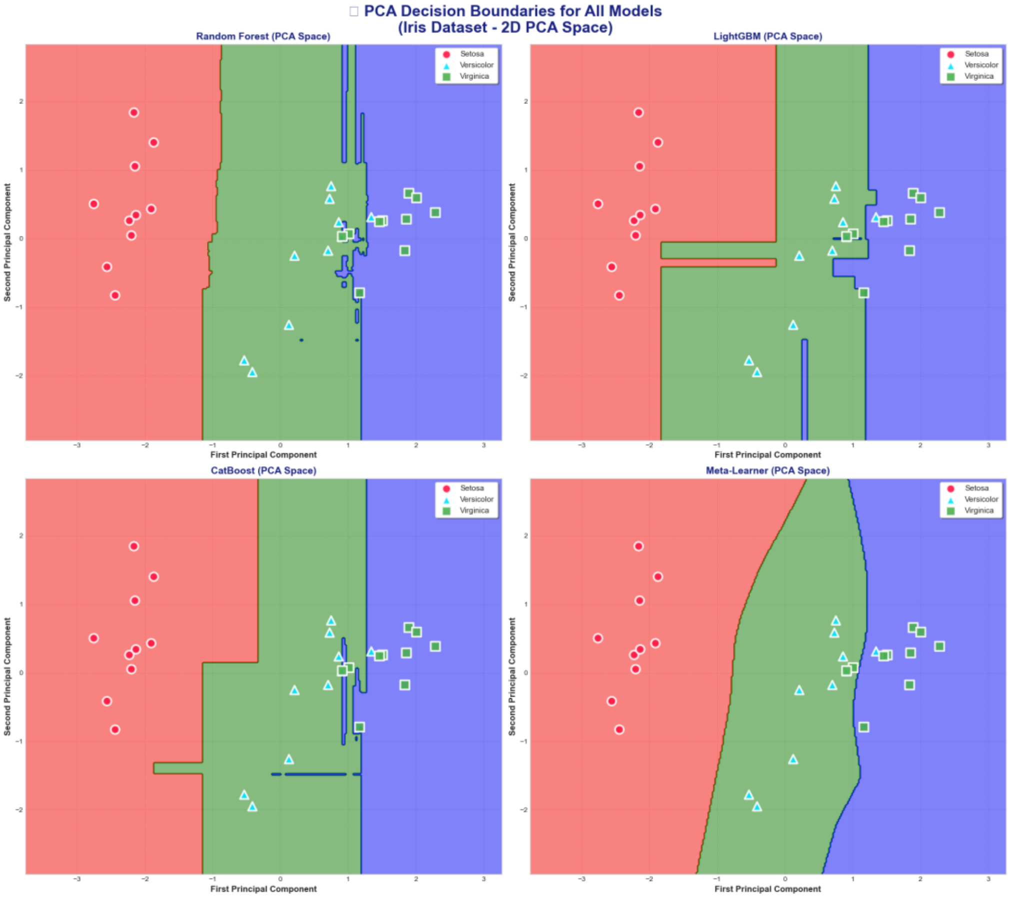

Fig. 14. Principal Component Analysis Decision Regions

While in Cross-Domain infrastructure PCA [16] in Fig. 15 reveals analogous patterns where Condition 1 (excellent) forms a distinct cluster separated from deteriorated states (Conditions 4-5), mirroring Setosa isolation. Infrastructure intermediate conditions (2-3) show greater overlap versus Iris Versicolor-Virginica, reflecting a gradual deterioration continuum. Meta-Feature PCA (infrastructure) demonstrates more compact clustering attributed to the dominant age predictor versus Iris's multi-dimensional morphology. Both visualizations validate classification feasibility in reduced dimensions, confirming that feature sets capture essential discriminative information.

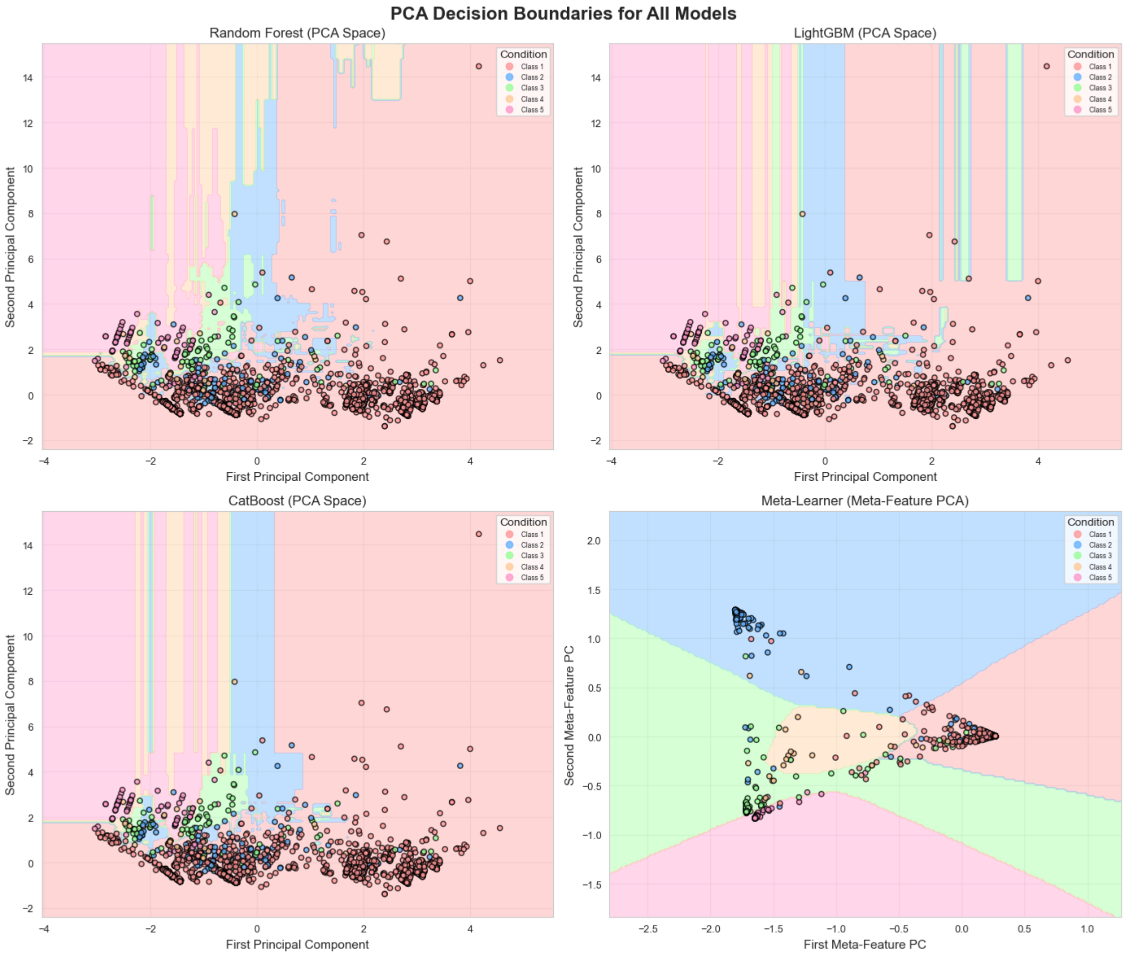

Fig. 15. PCA Decision Boundaries All Models in Cross-Domain Infrastructure [16]

Partial dependence plots in Fig. 16 reveal meta-model sensitivity to base learner predictions. Random Forest influence increases above a 0.6 probability threshold, demonstrating confidence-weighted integration. LightGBM shows the steepest response in the 0.4-0.8 range with maximum sensitivity at 0.6, indicating primary discriminative contribution. CatBoost maintains consistent influence across the full spectrum, suggesting stable supplementary pattern recognition. Vertical markers indicate data concentration regions validating statistical reliability. PDP analysis confirms meta-learner learns optimal weighting strategies rather than uniform averaging, enhancing ensemble intelligence. While in Cross-Domain, the infrastructure PDP [16] reveals analogous meta-learner dependence patterns where base learner probabilities differentially influence final predictions. Infrastructure demonstrates similar LightGBM prominence, reflecting gradient boosting's effectiveness across domains. However, the infrastructure's five-class problem [16] produces more complex dependence surfaces than Iris's three-class visualization. Both analyses validate meta-learning sophistication—ensemble doesn't simply average predictions but learns context-dependent weighting strategies. Universal pattern confirms meta-models extract algorithmic complementarity, justifying architectural complexity across scientific domains.

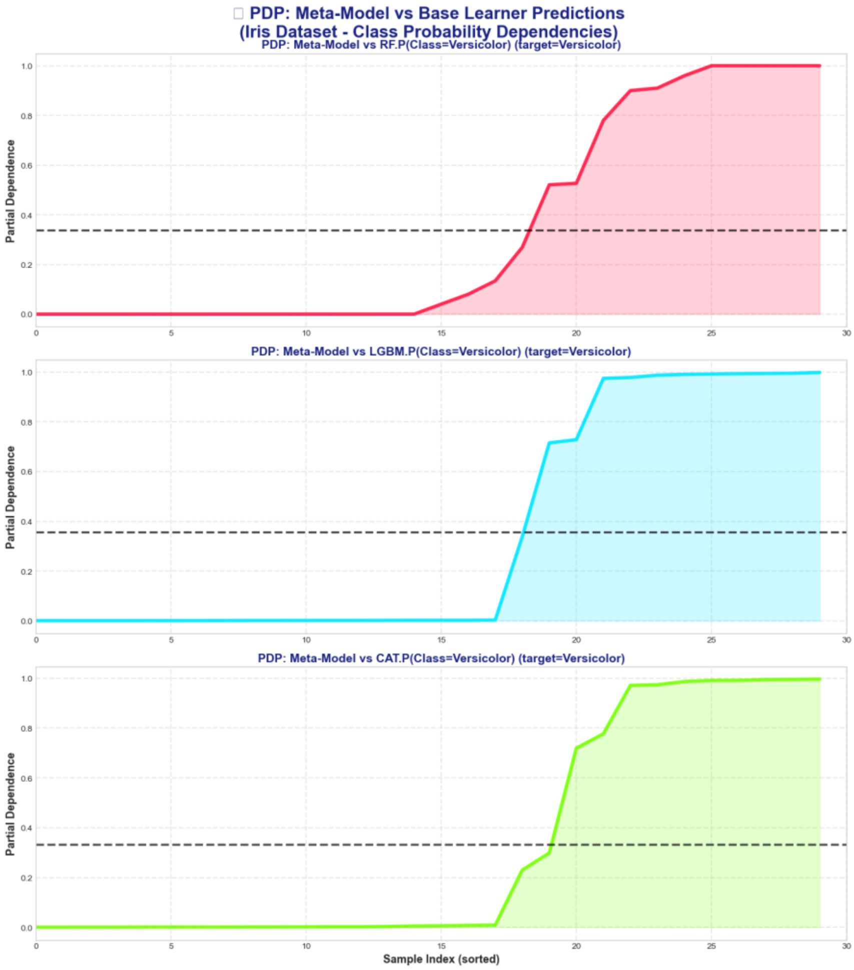

Fig. 16. Meta-Feature Contribution Analysis

Experimental evaluations demonstrated ensemble learning superiority across all performance metrics, with the meta-model achieving 93.33% accuracy through strategic base learner integration. Petal morphology emerged as the dominant discriminative feature, while cross-validation analysis validated prediction stability, establishing methodological foundations for subsequent interpretative discussion.

5. Discussion

Ensemble performance aligns with theoretical frameworks where Random Forest reduces variance through bagging methodology reduces variance through bootstrapped sampling [10], while gradient boosting algorithms (LightGBM, CatBoost) iteratively minimize prediction errors through sequential model refinement [24] [26]. Meta-model superiority validates stacking ensemble principles [12], where algorithmic diversity enables complementary pattern recognition beyond individual learner capabilities. Feature importance hierarchies corroborate botanical literature emphasizing reproductive morphology's taxonomic significance [6] [5], extending classical statistical foundations into contemporary machine learning frameworks.

Cross-domain methodology transferability demonstrates ensemble learning's universal applicability, paralleling infrastructure condition assessment findings [16], where meta-models achieved 96.67% accuracy through analogous stacking architectures. Performance differentials reflect dataset scale effects—infrastructure's 77-fold larger corpus facilitates superior generalization versus botanical classification's constrained sample size. Gradient boosting consistency across domains (LightGBM prominence) validates algorithmic robustness irrespective of application context, supporting ensemble method adoption for diverse classification problems requiring interpretability, accuracy, and operational reliability [51] [19].

6. Conclusion

This investigation established ensemble learning methodologies as frameworks for multi-class botanical classification by comprehensively benchmarking Random Forest, LightGBM, CatBoost, and stacking meta-models on the Iris dataset. Morphological feature hierarchies were quantitatively validated through multi-faceted importance analyses, confirming petal dimensions as primary discriminative characteristics. Meta-model architectures synthesized algorithmic diversity to achieve balanced classification performance, demonstrating transferable principles applicable across scientific domains, as evidenced through infrastructure assessment [16] parallels. These findings contribute methodological foundations for scalable, data-driven classification systems requiring domain-specific feature interaction modeling.

7. Recommendation

Future research should extend ensemble methodologies to larger botanical datasets exhibiting class imbalance and high-dimensional feature spaces, where advanced techniques [52] demonstrate measurable advantages. Neural Architecture Search (NAS) [53] could automate hyperparameter optimization [54] for domain-specific architectures [55], while deep ensemble methods [56] [57] may enhance uncertainty quantification for borderline classifications [58] [59]. Integration of explainable AI frameworks [60] would provide stakeholder-accessible interpretations beyond feature importance rankings [61]. Cross-domain transfer learning applications [62] could validate [63] the methodology's generalizability [64] across taxonomic hierarchies, agricultural phenotyping, and biodiversity monitoring systems that require real-time classification with limited training data.

Author Contributions

Conceptualization, M.M. and V.K.; methodology, M.M.; software, M.M..; validation, M.M.; formal analysis, M.M.; investigation, M.M.; resources, M.M.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M. and V.K.; visualization, M.M.; supervision, M.N., V.K. and A.J.; All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgement

This research was not funded by any grant.

References

[1] R. Gupta, A. Sharma, and T. Alam, "Building Predictive Models with Machine Learning," *Studies in Big Data*, vol. 145, pp. 39–59, 2024, doi: 10.1007/978-981-97-0448-4_3/FIGURES/3.

[2] I. H. Sarker, A. S. M. Kayes, and P. Watters, “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage,” *J Big Data*, vol. 6, no. 1, pp. 1–28, Dec. 2019, doi: 10.1186/S40537-019-0219-Y/FIGURES/5.

[3] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, “Machine Learning Interpretability: A Survey on Methods and Metrics,” *Electronics 2019, Vol. 8, Page 832*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/ELECTRONICS8080832.

[4] P. K. Kanti, P. Sharma, M. P. Maiya, and K. V. Sharma, “The stability and thermophysical properties of Al2O3-graphene oxide hybrid nanofluids for solar energy applications: Application of robust autoregressive modern machine learning technique,” *Solar Energy Materials and Solar Cells*, vol. 253, p. 112207, May 2023, doi: 10.1016/J.SOLMAT.2023.112207.

[5] A. Gorban and A. Zinovyev, “Principal Manifolds and Graphs in Practice,” *Int J Neural Syst*, vol. 20, no. 03, p. 219, Jan. 2010, [Online]. Available: http://arxiv.org/abs/1001.1122

[6] R. A. FISHER, “THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS,” *Ann Eugen*, vol. 7, no. 2, pp. 179–188, Sep. 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.

[7] S. Hossain *et al.*, “Classification of Iris Flower Dataset using Different Algorithms,” *International Journal of Scientific Research in Research Paper. Mathematical and Statistical Sciences*, vol. 9, no. 6, pp. 1–10, 2022, doi: 10.26438/ijsrmss/v9i6.110.

[8] K. J. Kim, H. Moskowitz, and M. Koksalan, “Fuzzy versus statistical linear regression,” *Eur J Oper Res*, vol. 92, no. 2, pp. 417–434, Jul. 1996, doi: 10.1016/0377-2217(94)00352-1.

[9] G. Sateesh Babu, P. Zhao, and X.-L. Li, “Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9642, Springer Verlag, 2016, pp. 214–228. doi: 10.1007/978-3-319-32025-0_14.

[10] L. Breiman, “Random forests,” *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.

[11] I. J. Bruno *et al.*, “New software for searching the Cambridge Structural Database and visualizing crystal structures,” *urn:issn:0108-7681*, vol. 58, no. 3, pp. 389–397, May 2002, doi: 10.1107/S0108768102003324.

[12] J. J. Zhu, M. Yang, and Z. J. Ren, “Machine Learning in Environmental Research: Common Pitfalls and Best Practices,” *Environ Sci Technol*, vol. 57, no. 46, pp. 17671–17689, Nov. 2023, doi: 10.1021/ACS.EST.3C00026/ASSET/IMAGES/LARGE/ES3C00026_0007.JPEG.

[13] B. Berlin, “Ethnobiological Classification,” *Cognition and Categorization*, pp. 9–26, Jan. 2024, doi: 10.4324/9781032633275-3/ETHNOBIOLOGICAL-CLASSIFICATION-BRENT-BERLIN.

[14] Z. Sun and T. Tan, “Ordinal measures for iris recognition,” *IEEE Trans Pattern Anal Mach Intell*, vol. 31, no. 12, pp. 2211–2226, 2009, doi: 10.1109/TPAMI.2008.240.

[15] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, “The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis,” *Lecture Notes in Networks and Systems*, vol. 750 LNNS, pp. 344–353, 2023, doi: 10.1007/978-3-031-42536-3_33/FIGURES/3.

[16] M. Mohammadagha, M. Najafi, V. Kaushal, and A. Jibreen, “Hybrid Machine Learning Meta-Model for the Condition Assessment of Urban Underground Pipes,” *Infrastructures (Basel)*, vol. 10, no. 11, p. 282, Oct. 2025, doi: 10.3390/INFRASTRUCTURES10110282.

[17] D. Khuong Nguyen, G. Sermpinis, C. Stasinakis, A. Smith Business School, and E. Financ, “Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology,” *European Financial Management*, vol. 29, no. 2, pp. 517–548, Mar. 2023, doi: 10.1111/EUFM.12365.

[18] S. Zhong *et al.*, “Machine Learning: New Ideas and Tools in Environmental Science and Engineering,” *Environ Sci Technol*, vol. 55, no. 19, pp. 12741–12754, Oct. 2021, doi: 10.1021/ACS.EST.1C01339/ASSET/IMAGES/LARGE/ES1C01339_0001.JPEG.

[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and Machine Learning forecasting methods: Concerns and ways forward,” *PLoS One*, vol. 13, no. 3, p. e0194889, Mar. 2018, doi: 10.1371/JOURNAL.PONE.0194889.

[20] H. Inam, N. U. Islam, M. U. Akram, and F. Ullah, “Smart and Automated Infrastructure Management: A Deep Learning Approach for Crack Detection in Bridge Images,” *Sustainability 2023, Vol. 15, Page 1866*, vol. 15, no. 3, p. 1866, Jan. 2023, doi: 10.3390/SU15031866.

[21] S. D. Datta, M. Islam, M. H. Rahman Sobuz, S. Ahmed, and M. Kar, “Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review,” *Heliyon*, vol. 10, no. 5, p. e26888, Mar. 2024, doi: 10.1016/J.HELIYON.2024.E26888/ASSET/3C33DBF1-44EB-4325-8187-2762E356BE7C/MAIN.ASSETS/GR4.JPG.

[22] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, “A theory of learning from different domains,” *Mach Learn*, vol. 79, no. 1–2, pp. 151–175, Oct. 2010, doi: 10.1007/S10994-009-5152-4/METRICS.

[23] B. Manavalan, S. Subramaniyam, T. H. Shin, M. O. Kim, and G. Lee, “Machine-Learning-Based Prediction of Cell-Penetrating Peptides and Their Uptake Efficiency with Improved Accuracy,” *J Proteome Res*, vol. 17, no. 8, pp. 2715–2726, Aug. 2018, doi: 10.1021/ACS.JPROTEOME.8B00148/ASSET/IMAGES/LARGE/PR-2018-00148H_0006.JPEG.

[24] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE.

[25] G. Ke *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” *Adv Neural Inf Process Syst*, vol. 30, 2017, Accessed: Nov. 30, 2025. [Online]. Available: https://github.com/Microsoft/LightGBM.

[26] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” *Adv Neural Inf Process Syst*, vol. 31, 2018, Accessed: Nov. 30, 2025. [Online]. Available: https://github.com/catboost/catboost

[27] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, “Hyperparameter Search for Machine Learning Algorithms for Optimizing the Computational Complexity,” *Processes 2023, Vol. 11, Page 349*, vol. 11, no. 2, p. 349, Jan. 2023, doi: 10.3390/PR11020349.

[28] K. Abbas *et al.*, “Measurements and determinants of extreme multidimensional energy poverty using machine learning,” *Energy*, vol. 251, p. 123977, Jul. 2022, doi: 10.1016/J.ENERGY.2022.123977.

[29] M. M. Giraldo-González and J. P. Rodríguez, “Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks,” *Water 2020, Vol. 12, Page 1153*, vol. 12, no. 4, p. 1153, Apr. 2020, doi: 10.3390/W12041153.

[30] S. S. Pakzad, N. Roshan, and M. Ghalehnovi, “Comparison of various machine learning algorithms used for compressive strength prediction of steel fiber-reinforced concrete,” *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–15, Mar. 2023, doi: 10.1038/s41598-023-30606-y.

[31] Y. J. Wong, S. K. Arumugasamy, C. H. Chung, A. Selvarajoo, and V. Sethu, “Comparative study of artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (Nephelium lappaceum) peel,” *Environ Monit Assess*, vol. 192, no. 7, pp. 1–20, Jun. 2020, doi: 10.1007/S10661-020-08268-4/TABLES/7.

[32] D. Kreuzberger, N. Kuhl, and S. Hirschl, “Machine Learning Operations (MLOps): Overview, Definition, and Architecture,” *IEEE Access*, vol. 11, pp. 31866–31879, 2023, doi: 10.1109/ACCESS.2023.3262138.

[33] Y. P. Zhang *et al.*, “Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling,” *Military Medical Research 2023 10:1*, vol. 10, no. 1, pp. 1–33, May 2023, doi: 10.1186/S40779-023-00458-8.

[34] M. W. Libbrecht and W. S. Noble, “Machine learning applications in genetics and genomics,” *Nature Reviews Genetics 2015 16:6*, vol. 16, no. 6, pp. 321–332, May 2015, doi: 10.1038/nrg3920.

[35] Y. Zhang, H. V. Burton, H. Sun, and M. Shokrabadi, “A machine learning framework for assessing post-earthquake structural safety,” *Structural Safety*, vol. 72, pp. 1–16, May 2018, doi: 10.1016/J.STRUSAFE.2017.12.001.

[36] N. Thapa, Z. Liu, D. B. Kc, B. Gokaraju, and K. Roy, “Comparison of Machine Learning and Deep Learning Models for Network Intrusion Detection Systems,” *Future Internet 2020, Vol. 12, Page 167*, vol. 12, no. 10, p. 167, Sep. 2020, doi: 10.3390/FI12100167.

[37] M. Der Yang and T. C. Su, “Automated diagnosis of sewer pipe defects based on machine learning approaches,” *Expert Syst Appl*, vol. 35, no. 3, pp. 1327–1337, Oct. 2008, doi: 10.1016/J.ESWA.2007.08.013.

[38] S. R. Shams *et al.*, “Assessing the effectiveness of artificial neural networks (ANN) and multiple linear regressions (MLR) in forcasting AQI and PM10 and evaluating health impacts through AirQ+ (case study: Tehran),” *Environmental Pollution*, vol. 338, p. 122623, Dec. 2023, doi: 10.1016/J.ENVPOL.2023.122623.

[39] G. Forkuor, O. K. L. Hounkpatin, G. Welp, and M. Thiel, “High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models,” *PLoS One*, vol. 12, no. 1, p. e0170478, Jan. 2017, doi: 10.1371/JOURNAL.PONE.0170478.

[40] J. A. J. Singh, J. Gnanasoundharam, M. Birunda, G. Sudha, S. P. Maniraj, and C. Srinivasan, “Wearable Sepsis Early Warning Using Cloud Computing and Logistic Regression Predictive Analytics,” *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2024*, 2024, doi: 10.1109/ICRITO61523.2024.10522250.

[41] Y. N. TAHIR, U. AMINU, H. S. DJINGARE, and A. S. AJIBOLA, “UTILIZING ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS) FOR COMPLEX PATTERN RECOGNITION AND SPECIES PREDICTION FOR IRIS DATASET,” *International Journal of Engineering Processing and Safety Research*, vol. 5, no. 5, Sep. 2024, Accessed: Mar. 22, 2025. [Online]. Available: https://cambridgeresearchpub.com/ijepsr/article/view/312

[42] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. M. Nagem, “A multi-biometric iris recognition system based on a deep learning approach,” *Pattern Analysis and Applications*, vol. 21, no. 3, pp. 783–802, Aug. 2018, doi: 10.1007/s10044-017-0656-1.

[43] Alex Dillhoff, “Neural Network Library,” 2025. [Online]. Available: https://ajdillhoff.github.io/

[44] R. Fisher, “Iris - UCI Machine Learning Repository,” 1936. doi: doi.org/10.24432/C56C76.

[45] A. Zollanvari, “Supervised Learning in Practice: the First Application Using Scikit-Learn,” *Machine Learning with Python*, pp. 111–131, 2023, doi: 10.1007/978-3-031-33342-2_4.

[46] L. Liu and R. C. Chen, “A novel passenger flow prediction model using deep learning methods,” *Transp Res Part C Emerg Technol*, vol. 84, pp. 74–91, Nov. 2017, doi: 10.1016/J.TRC.2017.08.001.

[47] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock market index using fusion of machine learning techniques,” *Expert Syst Appl*, vol. 42, no. 4, pp. 2162–2172, Mar. 2015, doi: 10.1016/J.ESWA.2014.10.031.

[48] Ronald Fisher, “Iris flower data set - Wikipedia,” 1936. [Online]. Available: https://en.wikipedia.org/wiki/Iris_flower_data_set

[49] Nicoguaro, “Iris dataset scatterplot.svg - Wikimedia Commons,” Jan. 09, 2016. [Online]. Available: https://commons.wikimedia.org/w/index.php?curid=46257808

[50] A. A. Ibrahim, R. L. Ridwan, M. M. Muhammed, R. O. Abdulaziz, and G. A. Saheed, “Comparison of the CatBoost Classifier with other Machine Learning Methods,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 738–748, Dec. 2020, doi: 10.14569/IJACSA.2020.0111190.

[51] T. G. Dietterich, “Ensemble Methods in Machine Learning,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1.

[52] Y. Zhou *et al.*, “Mixture-of-Experts with Expert Choice Routing,” *Adv Neural Inf Process Syst*, vol. 35, pp. 7103–7114, Dec. 2022.

[53] B. Zoph and Q. V. Le, “Neural Architecture Search with Reinforcement Learning,” *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, Nov. 2016, Accessed: Jun. 21, 2025. [Online]. Available: https://arxiv.org/pdf/1611.01578

[54] J. Thiyagalingam, M. Shankar, G. Fox, and T. Hey, “Scientific machine learning benchmarks,” *Nature Reviews Physics*, vol. 4, no. 6, pp. 413–420, Jun. 2022, doi: 10.1038/S42254-022-00441-7;SUBJMETA=1046,172,639,704,705;KWRD=ENVIRONMENTAL+SCIENCES,SCIENTIFIC+DATA.

[55] R. Maulik, O. San, J. D. Jacob, and C. Crick, “Sub-grid scale model classification and blending through deep learning,” *J Fluid Mech*, vol. 870, pp. 784–812, Jul. 2019, doi: 10.1017/JFM.2019.254.

[56] B. Pavlyshenko, “Using Stacking Approaches for Machine Learning Models,” *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP 2018*, pp. 255–258, Oct. 2018, doi: 10.1109/DSMP.2018.8478522.

[57] A. Mosavi, F. Sajedi Hosseini, B. Choubin, M. Goodarzi, A. A. Dineva, and E. Rafiei Sardooi, “Ensemble Boosting and Bagging Based Machine Learning Models for Groundwater Potential Prediction,” *Water Resources Management*, vol. 35, no. 1, pp. 23–37, Jan. 2021, doi: 10.1007/S11269-020-02704-3/TABLES/3.

[58] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, “PointFlow: 3D Point Cloud Generation With Continuous Normalizing Flows,” 2019.

[59] E. Bauer and R. Kohavi, “Empirical comparison of voting classification algorithms: bagging, boosting, and variants,” *Mach Learn*, vol. 36, no. 1, pp. 105–139, 1999, doi: 10.1023/A:1007515423169/METRICS.

[60] E. Prem, “From ethical AI frameworks to tools: a review of approaches,” *AI and Ethics*, vol. 3, no. 3, pp. 699–716, Aug. 2023, doi: 10.1007/s43681-023-00258-9.

[61] N. Du *et al.*, “GLaM: Efficient Scaling of Language Models with Mixture-of-Experts,” PMLR, Jun. 2022. Accessed: Jun. 21, 2025. [Online]. Available: https://proceedings.mlr.press/v162/du22c.html

[62] M. Abdin *et al.*, “Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone,” Apr. 2024, Accessed: Jun. 21, 2025. [Online]. Available: https://arxiv.org/pdf/2404.14219

[63] J. Liu *et al.*, “Adaptive Asynchronous Federated Learning in Resource-Constrained Edge Computing,” *IEEE Trans Mob Comput*, vol. 22, no. 2, pp. 674–690, Feb. 2023, doi: 10.1109/TMC.2021.3096846.

[64] A. Wayne McCord and W. Creek, “Method and system for enabling automated and real-time discovery of skills available to agents and systems in a multimedia communications network,” Jun. 2003, Accessed: Jun. 21, 2025. [Online]. Available: http://computer.org/internet?,