MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2 (2026) | Pages: 487-493

An Intelligent Machine Learning Framework for Predictive Analytics and Decision-Making in Computer Science Applications

Dipannita Mondal*

Department of Artificial Intelligence- Data Science and Machine Learning, Dr. D. Y. Patil College of Engineering and Innovation Talegaon, Varale, Pune, Maharashtra, India

**Corresponding Author: ddmondal2684@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 21 July 2026*
***Revised**: 13 Aug 2026*
***Accepted**: 19 Sept 2026*
***Published**: 14 Oct 2026*

Abstract

The rapid expansion of digital technologies, distributed computing environments, and data-intensive applications has increased the demand for intelligent predictive systems capable of supporting accurate, efficient, and interpretable decision-making. Traditional computational approaches frequently experience limitations in processing heterogeneous datasets, identifying nonlinear relationships, and adapting to changing operational conditions. This study proposes an intelligent machine learning framework integrating data preprocessing, feature engineering, ensemble predictive modeling, probability calibration, explainable artificial intelligence, and cost-sensitive decision optimization. The proposed architecture combines Random Forest, Extreme Gradient Boosting, and Multilayer Perceptron learning to generate complementary predictions, which are integrated through validation-based weighted ensemble techniques. An explainability mechanism employing Shapley Additive explanations supports the interpretation of model outputs, while an optimization component transforms calibrated predictions into application-specific decisions. The framework is designed for computer science applications including cybersecurity intrusion detection, software defect prediction, cloud resource management, and computational performance monitoring. A reproducible experimental protocol is formulated to evaluate classification accuracy, precision, recall, F1-score, probability calibration, computational overhead, and decision cost. Recent research published through provides the theoretical foundation for combining ensemble learning, explainable prediction, and risk-aware decision support. The proposed framework establishes a methodological basis for developing intelligent systems that connect predictive analysis with transparent operational decisions. Empirical validation using benchmark datasets is identified as the next essential stage for determining its comparative effectiveness.

Keywords: Machine Learning; Predictive Analytics; Decision-Making; Ensemble Learning; Explainable Artificial Intelligence; Intelligent Systems

How to Cite This Article

Mondal, D. (2026). An intelligent machine learning framework for predictive analytics and decision-making in computer science applications. *Multidisciplinary Journal of Research in Engineering and Technology,* 13(2), 487–493.

2348-6953/ © 2026 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0).

Introduction

The increasing integration of artificial intelligence into modern computing environments has transformed how organizations process information, predict operational events, and formulate decisions. Cloud computing platforms, cybersecurity monitoring systems, software engineering repositories, distributed networks, and enterprise information systems continuously generate extensive volumes of heterogeneous data. These datasets often contain valuable patterns that can support the early identification of operational problems, resource utilization forecasting, software failure detection, and intelligent system management. Nevertheless, converting such information into reliable predictions and actionable knowledge remains challenging because of high-dimensional feature spaces, incomplete observations, nonlinear dependencies, and constantly evolving computational environments.

Machine learning has become a significant component of predictive analytics because it enables computational models to discover patterns and statistical relationships from historical observations. Unlike traditional rule-based systems, machine learning algorithms can adapt their predictive functions to complex data distributions. Supervised learning algorithms are widely employed for classification and regression problems, while ensemble learning integrates several predictive models to improve robustness and generalization. Recent research demonstrates that tree-based methods, neural architectures, and heterogeneous ensembles remain important techniques for structured-data prediction, although their comparative effectiveness depends substantially on datasets, validation methods, and computational budgets (Erickson et al., 2025).

Predictive analytics has relevance in computer science applications. Within cybersecurity, learning algorithms can identify potentially malicious network activities through unusual traffic patterns. Software engineering systems can estimate defect risk using code complexity, historical failures, and development metrics. Cloud computing platforms can employ predictive approaches to estimate resource demand, identify abnormal system utilization, and support proactive resource management. These applications illustrate how predictive information can contribute to operational improvement when combined with appropriate decision mechanisms.

Despite considerable advances in machine learning algorithms, several challenges continue to restrict effective implementation. First, the predictive performance of a single algorithm may fluctuate across datasets containing different levels of noise, feature dependencies, and class imbalance. Second, highly complex models can generate outputs that are difficult for users to interpret, potentially reducing confidence in automated decision-support systems. Third, predicted class labels or probability estimates do not inherently identify the most appropriate operational response. Accurate decisions depend not only on identifying likely outcomes but also on understanding the consequences of alternative actions.

Recent studies emphasize the importance of interpretability and uncertainty-aware decision support. Research on explainable artificial intelligence indicates that transparent explanations can improve the examination and communication of algorithmic decisions, although explanation reliability must be evaluated separately from prediction accuracy. Similarly, cost-aware calibration and cost-sensitive classification studies demonstrate that prediction probabilities and decision consequences must be considered together when developing practical machine learning applications (Yang & Bi, 2024; Komisarenko & Kull, 2025).

In response to these limitations, the present study proposes an Intelligent Machine Learning Framework for Predictive Analytics and Decision-Making (IML-PADM). The framework integrates multiple learning algorithms, adaptive feature processing, validation-based prediction aggregation, probability calibration, explainable artificial intelligence, and cost-sensitive action optimization within modular architecture. Its central purpose is to support computer science applications that require predictions to be converted into understandable and operationally meaningful decisions.

The proposed research adopts a design-oriented methodology and establishes an experimental protocol for subsequent implementation. Unlike approaches that focus solely on improving classification metrics, the framework evaluates predictive performance together with decision cost, computational efficiency, and interpretability. This integrated perspective provides a foundation for investigating the practical relationship between statistical learning and intelligent decision support.

Literature Review

Fan et al. (2025) presented a comprehensive survey examining ensemble diversity, advanced model combinations, high-dimensional learning, and application challenges. Their analysis supports the principle that heterogeneous models can offer complementary predictive characteristics, but the benefits of combining them depend on appropriately managing diversity and overfitting.

Comparative evaluation of machine learning algorithms remains an active research area. Pasaribu et al. (2024) investigated XGBoost alongside tabular deep learning techniques and demonstrated that the relative effectiveness of algorithms differs between classification and regression settings. The findings indicate that no individual architecture should automatically be treated as universally superior. Consequently, systematic validation and task-sensitive algorithm selection are required when designing generalized predictive frameworks.

Erickson et al. (2025) introduced TabArena, a continuously maintained benchmark for tabular machine learning. Their work emphasizes the importance of transparent evaluation protocols, consistent model implementations, and careful comparison of ensemble learning strategies. It also highlights the potential for validation-set overfitting when multiple configurations and ensemble combinations are repeatedly evaluated. These observations are particularly relevant to the proposed framework because ensemble integration requires independent validation and an untouched test dataset.

Explainability is another important dimension of intelligent predictive systems. Research published in Decision Support Systems in 2024 examined how explainable artificial intelligence can contribute to organizational decision-making through data, methodological, and application-level perspectives. This research reinforces the need to consider the user's understanding of model behavior rather than treating predictive accuracy as the only evaluation criterion.

Löfström et al. (2024) investigated calibrated explanations that incorporate uncertainty information and counterfactual reasoning. Their work addresses limitations of conventional feature importance explanations, particularly instability and uncertainty. Such findings motivate the inclusion of model calibration and explanation analysis within a common decision-support architecture.

Löfström et al. (2025) subsequently examined calibrated explanations for regression, extending uncertainty-aware interpretability to continuous prediction tasks. Their research is important for computer science applications involving resource demand, system latency, or computational load, where predicted outcomes are often numerical rather than categorical.

Kopanja et al. (2024) proposed cost-sensitive tree SHAP approaches for interpreting cost-sensitive tree-based models. Their work reflects the growing recognition that explanatory mechanisms should account for practical decision considerations. Explanation relevance may depend on the objectives and costs associated with the model's application rather than feature contributions alone.

Research on calibration has also emphasized the relationship between probability estimation and decision consequences. Yang and Bi (2024) examined cost-aware calibration and established methods for considering asymmetric costs associated with inaccurate probability estimates. Their investigation provides a direct theoretical foundation for integrating calibrated predictions with downstream decision policies.

Komisarenko and Kull (2025) investigated classification under uncertainty regarding the costs of prediction errors. Their research evaluated cost-sensitive loss functions and examined how standard learning objectives interact with operational decision requirements. The study demonstrates the importance of considering uncertain cost structures when interpreting model performance.

Methodology

Research Design

The proposed research adopts a quantitative, design-oriented methodology for developing a machine learning-based predictive and decision-support architecture. The framework is organized into six interconnected stages: data collection, preprocessing, feature engineering, predictive model development, explainability and probability calibration, and intelligent decision optimization. This layered design is intended to facilitate implementation across heterogeneous computer science applications while maintaining a consistent analytical structure.

The dataset is mathematically represented as:

$$D = \{(X_i, Y_i)\}_{i=1}^{N} \quad \text{-----(1)}$$

where (D) represents the complete dataset, (X_i) denotes the input feature vector, (Y_i) represents the target output, and (N) indicates the total number of observations. The framework learns a predictive function (f(X)) capable of estimating the relationship between observed system characteristics and the corresponding outcomes. The resulting predictions are subsequently processed through calibration, interpretability, and decision-selection mechanisms.

Proposed Framework Architecture

The architecture integrates data-driven modeling and decision intelligence. Its modular arrangement enables each component to be modified independently, depending on the computational problem under investigation.

Figure 1 illustrates the proposed intelligent machine learning framework integrating eight interconnected stages: data sources, data acquisition and integration, data preprocessing, feature engineering and selection, intelligent machine learning, predictive analytics, decision-making support, and visualization. The framework processes structured and unstructured data to generate accurate predictions and support informed decision-making. A continuous feedback mechanism enables model updates and performance improvement, while security, scalability, and explainability support all stages. The architecture provides a conceptual foundation for developing adaptive, reliable, and interpretable predictive analytics systems across diverse computer science applications.

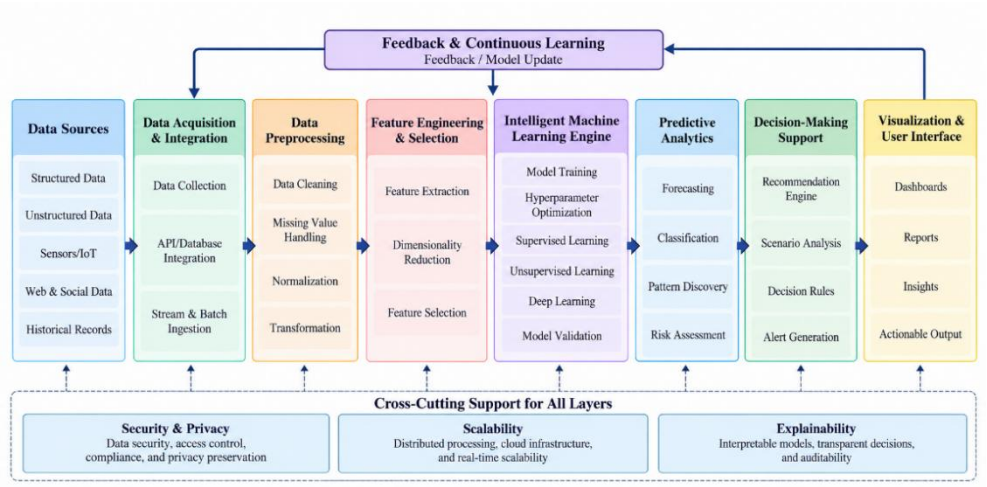


Fig. 1. *Proposed Architecture of the Intelligent Machine Learning Framework for Predictive Analytics and Decision-Making*

Feature Engineering and Selection

Feature engineering transforms raw observations into meaningful predictive variables. It is particularly important in computational applications where system behavior may depend on interactions among network activity, resource consumption, execution patterns, and software characteristics.

The framework considers statistical transformations, domain-specific feature construction, correlation analysis, and model-based importance estimation. Selection procedures must be performed exclusively within the training process to prevent data leakage.

The feature transformation process is expressed as:

$$Z_i = \phi(X_i; \theta) \quad \text{--------(2)}$$

where (Zi) represents the transformed feature vector, (phi) denotes the preprocessing and feature engineering operation, and (theta) represents parameters determined from the training data.

Weighted Ensemble Prediction

The framework integrates the predicted probabilities generated by constituent models. The ensemble output for class (c) is calculated as:

$$P_E(c \mid X) = \sum_{j=1}^{M} w_j \, P_j(c \mid X) \quad \text{------(3)}$$

Subject to:

$$\sum_{j=1}^{M} w_j = 1, \quad w_j \geq 0 \quad \text{-----(4)}$$

where (M) represents the number of models, (Pj(cmid X)) denotes the probability produced by model (j), and (wj) indicates its contribution to the final prediction.

Ensemble weights are determined using out-of-fold or appropriately held-out validation predictions. Optimization may minimize cross-entropy or another proper scoring rule, subject to nonnegative weight constraints. This procedure is designed to prevent direct optimization against the final testing dataset.

Probability Calibration and Explainability

The predictive output is passed through a calibration component intended to improve the reliability of probability estimates. Calibration is important in applications where action thresholds depend upon estimated risk. Suitable techniques include sigmoid calibration and isotonic regression, selected according to the characteristics of the calibration dataset.

The framework incorporates SHAP-based explanations to identify the features contributing to model predictions. For a prediction function represented through an additive explanation:

$$f(X) = \phi_0 + \sum_{k=1}^{d} \phi_k \quad \text{----(5)}$$

where (phi_0) denotes the baseline prediction and (phi_k) represents the contribution assigned to feature (k). The exact interpretation depends on the explanation method and output scale.

Experimental Design and Algorithm

Proposed Implementation Configuration

Table 1. *Proposed experimental configuration*

Parameter	Proposed specification
Programming language	Python
Model development	scikit-learn, XGBoost
Data processing	pandas, NumPy
Base models	Random Forest, XGBoost, MLP
Baselines	Logistic Regression, Decision Tree
Dataset partition	70:15:15, where appropriate
Cross-validation	Stratified five-fold
Ensemble strategy	Validation-optimized weighted averaging
Calibration	Sigmoid or isotonic
Explainability	SHAP
Primary metrics	F1-score, ROC-AUC, Brier score
Operational metrics	Prediction latency, decision cost

All entries describe a proposed setup, not completed experimental activities.

Experimental Evaluation Metrics

The predictive models should be compared using classification and operational performance metrics. Accuracy assesses the overall percentage of correct classifications, while precision and recall provide additional information concerning false positives and false negatives.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{-----(6)}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{-------(7)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{-------(8)}$$

$$F1 = 2\frac{Precision \times Recall}{Precision+Recall} \quad \text{-------(9)}$$

Where (TP), (TN), (FP), and (FN) represent true positives, true negatives, false positives, and false negatives, respectively.

The Brier score is included for binary probability evaluation:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \quad \text{------(10)}$$

where (pi) is the predicted probability of the positive class and (yi) is its observed indicator. Lower values indicate better probabilistic accuracy.

Decision quality should be measured using realized operational cost on held-out data. For example, a system that generates fewer costly missed detections may be preferable even when its overall accuracy is comparable to another classifier.

Results and Discussion

Proposed Comparative Evaluation

The framework has been formulated to support comparisons between conventional machine learning algorithms and integrated predictive decision-support configurations. Evaluation should distinguish improvements in predictive discrimination from improvements associated with probability calibration and decision optimization.

Table 2. *Comparative results reporting template*

Model	Accuracy	F1-score	ROC-AUC	Brier score
Logistic Regression	Pending	Pending	Pending	Pending

Decision Tree	Pending	Pending	Pending	Pending
Random Forest	Pending	Pending	Pending	Pending
XGBoost	Pending	Pending	Pending	Pending
MLP	Pending	Pending	Pending	Pending
Weighted Ensemble	Pending	Pending	Pending	Pending
Full Proposed Framework	Pending	Pending	Pending	Pending

The full proposed framework and weighted ensemble may produce the same class predictions, depending on the calibration and decision configurations. Therefore, improvement in operational value should not be represented as an automatic improvement in classification accuracy.

Discussion of Predictive Performance

The principal motivation for using heterogeneous ensemble learning is that models with different inductive biases may capture complementary patterns. Random Forest can offer robustness against some forms of noise, while boosting can identify detailed nonlinear relationships. Neural models may learn flexible representations when suitable feature structure and training data are available. Recent ensemble-learning research supports evaluating model diversity, but it also identifies limitations associated with validation overfitting and increased computational requirements. Consequently, the effectiveness of the proposed architecture depends on appropriate dataset selection, independent testing, and control of model complexity. A further consideration is the difference between classification confidence and actual outcome probability. Systems that generate poorly calibrated confidence estimates may make undesirable risk-sensitive decisions even when their class predictions are frequently correct. The inclusion of a calibration module is therefore intended to improve the usefulness of predictions for downstream optimization rather than guarantee an improvement in all classification metrics.

Conclusion

This study proposes an intelligent machine learning framework for predictive analytics and decision-making in computer science applications. The architecture integrates data preprocessing, adaptive feature engineering, ensemble-based prediction, probability calibration, explainable artificial intelligence, and cost-sensitive decision optimization within a modular analytical system. The framework is intended to address important limitations of predictive models that focus primarily on classification accuracy without considering the interpretation, reliability, and operational consequences of their outputs. By combining Random Forest, XGBoost, and Multilayer Perceptron learning, the proposed architecture provides a structured environment for investigating heterogeneous prediction strategies. The introduction of probability calibration and SHAP-based explanations extends the system beyond conventional predictive modeling, while the decision optimization module connects estimated outcomes with application-specific costs. Recent literature published through 2025 emphasizes the continuing importance of ensemble learning, reliable benchmarking, uncertainty-aware explanations, and cost-sensitive decision analysis. These developments provide a theoretical foundation for the proposed framework and support its relevance to contemporary computer science research. The primary contribution of the study is a unified methodological design that enables predictive performance and decision usefulness to be evaluated within a consistent architecture. Its applicability to cybersecurity monitoring, software defect prediction, cloud resource optimization, and other data-intensive computing domains establishes a basis for future cross-domain research. Future investigations should implement the framework using publicly available benchmark datasets and evaluate its performance against strong baseline methods. Additional research may incorporate concept-drift detection, online learning, explainability stability assessment, privacy-preserving analytics, and real-time optimization. Rigorous empirical validation will be necessary to determine whether the integrated architecture offers measurable benefits under realistic computational and operational conditions.

References

1. Erickson, N., Purucker, L., Tschalzev, A., Holzmüller, D., Desai, P., Salinas, D., & Hutter, F. (2025). TabArena: A living benchmark for machine learning on tabular data. Advances in Neural Information Processing Systems, 38. https://doi.org/10.52202/085713-0519
2. Fan, Z., Yu, Z., Yang, K., Chen, W., Liu, X., Li, G., Yang, X., & Chen, C. L. P. (2025). Diverse models, united goal: A comprehensive survey of ensemble learning. CAAI Transactions on Intelligence Technology, 10(4), 959–982. https://doi.org/10.1049/cit2.70030
3. Komisarenko, V., & Kull, M. (2025). Cost-sensitive classification with cost uncertainty: Do we need surrogate losses? Machine Learning, 114, Article 132. https://doi.org/10.1007/s10994-024-06634-8
4. Löfström, T., Löfström, H., Johansson, U., Sönströd, C., & Matela, R. (2025). Calibrated explanations for regression. Machine Learning, 114, Article 100. https://doi.org/10.1007/s10994-024-06642-8

5. Yang, M., & Bi, X. (2024). Cost-aware calibration of classifiers. INFORMS Journal on Data Science, 4(2), 101–113. https://doi.org/10.1287/ijds.2024.0038
6. Kopanja, M., Hačko, S., Brdar, S., & Savić, M. (2024). Cost-sensitive tree SHAP for explaining cost-sensitive tree-based models. Computational Intelligence, 40(3), Article e12651. https://doi.org/10.1111/coin.12651
7. Löfström, H., Löfström, T., Johansson, U., & Sönströd, C. (2024). Calibrated explanations: With uncertainty information and counterfactuals. Expert Systems with Applications, 246, Article 123154. https://doi.org/10.1016/j.eswa.2024.123154
8. Pasaribu, J., Yudistira, N., & Mahmudy, W. F. (2024). Tabular data classification and regression: XGBoost or deep learning with retrieval-augmented generation. IEEE Access, 12, 191719–191732. https://doi.org/10.1109/ACCESS.2024.3518205
9. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785
11. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. Proceedings of Machine Learning Research, 70, 1321–1330. https://proceedings.mlr.press/v70/guo17a.html
12. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. https://proceedings.neurips.cc/paper/2017
13. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778