Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian and Machine Learning Approaches to the Classification Problem

Romuald Daniel BOY-NGBOGBELE

romuald.boy-ngbogbele@students.jkuat.ac.ke

Pan African University, Nairobi

Research Article

Keywords: Imbalanced Classification, Bayesian Inference, Machine Learning, Hybrid Ensemble Methods, Probabilistic Modeling

Posted Date: May 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9621535/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Bayesian and Machine Learning Approaches to the Classification Problem

Romuald Daniel Boy-ngbogbele *[1]

[1]Department of Mathematics, Pan African University Institute for Basic Science Technology and Innovation, Nairobi, Kenya

Abstract

Background: Classification problems are fundamental in statistical learning and arise in numerous fields such as healthcare, finance, and environmental sciences. Traditional statistical models provide interpretable parameter estimates but may struggle to capture complex nonlinear relationships, whereas machine learning methods often achieve strong predictive performance at the cost of interpretability. These limitations motivate the development of hybrid approaches that integrate probabilistic modeling with modern machine learning techniques.

Methods: This study proposes a Bayesian–Machine Learning Ensemble (BMLE) framework that combines Bayesian logistic regression with a machine learning component to improve classification performance while maintaining interpretability. A simulation study was conducted to evaluate the proposed model and compare its performance with a conventional logistic regression model. Model performance was assessed using several classification metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). In addition, parameter estimation was evaluated using bias, root mean squared error (RMSE), and coverage probability.

Results: The simulation results demonstrate that the proposed BMLE model consistently outperforms the conventional logistic regression model across all predictive performance metrics. The corrected model achieved higher accuracy, precision, recall, and F1-score, as well as a higher AUC value, indicating improved discriminative ability. Furthermore, the Bayesian component provided stable parameter estimates and meaningful uncertainty quantification through credible intervals.

Conclusion: The proposed Bayesian–Machine Learning Ensemble framework offers a flexible and robust approach for classification problems by integrating probabilistic inference with machine learning techniques. The results suggest that the hybrid modeling strategy improves predictive accuracy while preserving interpretability, making it a promising methodology for complex classification tasks.

Keywords: Imbalanced Classification; Bayesian Inference; Machine Learning; Hybrid Ensemble Methods; Probabilistic Modeling.

1 Introduction

Classification is one of the most fundamental tasks in modern data analysis and supervised learning. It involves assigning observations to predefined categories based on a set of predictor variables. Classification methods are widely applied in diverse domains such as medical diagnosis, financial risk assessment,

*Corresponding author: R. D. B. romuald.boy-ngbogbele@students.jkuat.ac.ke

environmental monitoring, and social sciences. With the rapid expansion of digital data and computational capabilities, classification models have become essential tools for extracting actionable insights from complex datasets. However, the increasing scale, dimensionality, and heterogeneity of modern data present substantial challenges for traditional statistical modeling approaches (Zhang et al., 2023; Kotu and Deshpande, 2023).

Classical statistical methods such as logistic regression and discriminant analysis have long served as foundational tools for classification tasks due to their interpretability and well-understood statistical properties. These models allow researchers to quantify relationships between predictors and outcomes while providing interpretable parameter estimates. Nevertheless, traditional models often rely on restrictive assumptions regarding linearity and independence among predictors. Such assumptions may limit their applicability when analyzing high-dimensional or highly nonlinear datasets that are increasingly common in contemporary data science applications (James et al., 2023).

Recent developments in machine learning have significantly expanded the range of tools available for classification problems. Algorithms such as random forests, gradient boosting machines, and deep neural networks have demonstrated remarkable predictive performance across many fields. These methods are particularly effective in capturing complex nonlinear relationships and interactions between variables without requiring strong parametric assumptions. Their scalability and ability to process large datasets have made them central to modern predictive analytics (Murphy, 2023; Goodfellow et al., 2024).

Despite their predictive power, many machine learning models are often criticized for their lack of interpretability. In high-stakes domains such as healthcare, finance, and public policy, decision makers frequently require transparent models that provide clear explanations of predictions and quantification of uncertainty. Black-box models may hinder trust and limit their practical adoption, particularly when accountability and regulatory requirements demand interpretable decision-making frameworks (Molnar, 2024).

Bayesian statistical methods offer a coherent probabilistic framework that naturally incorporates uncertainty into the modeling process. By combining prior knowledge with observed data, Bayesian approaches allow researchers to obtain full posterior distributions for model parameters and predictions. This capability enables more comprehensive uncertainty quantification and facilitates principled decision-making under uncertainty. Bayesian models are therefore particularly attractive in contexts where prior information is available or where uncertainty plays a critical role in interpretation (Gelman et al., 2024).

In addition to uncertainty quantification, Bayesian modeling provides flexible hierarchical structures that allow information to be shared across related groups or datasets. Such hierarchical models are especially useful when dealing with heterogeneous populations, sparse observations, or rare events. They can improve parameter estimation and predictive performance by borrowing strength across groups, thereby stabilizing inference in complex datasets (McElreath, 2024).

Another major challenge frequently encountered in classification tasks is class imbalance. In many real-world applications, the events of greatest interest correspond to rare outcomes, such as fraudulent transactions, equipment failures, or the presence of disease. Traditional classification algorithms often perform poorly in such scenarios because they tend to favor the majority class, leading to reduced sensitivity to minority events. Addressing class imbalance remains a critical issue in the design of reliable predictive models (Johnson and Khoshgoftaar, 2023).

Recent research suggests that combining Bayesian inference with modern machine learning techniques may offer a promising strategy for addressing the limitations of each paradigm individually. Machine learning models provide strong predictive capabilities and scalability, while Bayesian methods contribute principled uncertainty quantification and interpretability. Hybrid frameworks that integrate

these approaches have the potential to produce models that are both accurate and transparent (Nguyen et al., 2024).

The rapid growth of large-scale datasets and high-dimensional feature spaces further motivates the development of classification models that can effectively balance predictive performance, interpretability, and computational efficiency. Advances in probabilistic programming, scalable inference algorithms, and deep learning architectures have created new opportunities for integrating statistical and algorithmic approaches to classification problems (Blei et al., 2023).

The objective of this study is therefore to investigate and develop classification frameworks that combine Bayesian statistical modeling with modern machine learning techniques. Specifically, the research aims to design methods capable of handling high-dimensional data, addressing class imbalance, and providing interpretable probabilistic predictions. By integrating these complementary paradigms, the study seeks to contribute to the development of robust, scalable, and transparent classification models applicable to a wide range of real-world data analysis problems.

2 Materials and Methods

2.1 Study Design

This study adopts a quantitative methodological research design aimed at developing and evaluating a hybrid Bayesian–machine learning framework for classification problems. The design integrates probabilistic modeling, machine learning algorithms, simulation experiments, and empirical evaluation. The study consists of three major components: (i) development of a Bayesian classification model capable of representing uncertainty through posterior distributions, (ii) implementation of machine learning algorithms for predictive learning from complex data structures, and (iii) construction of a hybrid Bayesian–Machine Learning ensemble model that integrates the strengths of both approaches.

To ensure reproducibility and rigorous evaluation, the study follows a structured workflow including data preparation, model specification, parameter estimation, model training, and performance evaluation. Existing statistical and machine learning methodologies are used as reference baselines and compared with the proposed hybrid model.

2.2 Study Setting

The methodological framework is implemented using both simulated datasets and publicly available real-world datasets. Simulated datasets are generated to systematically control key characteristics such as sample size, dimensionality, and class imbalance. These controlled experiments allow the performance of the proposed model to be evaluated under different data-generating scenarios.

In addition to simulated data, publicly available datasets from open repositories such as the UCI Machine Learning Repository and Kaggle are used to evaluate the external validity of the proposed model. These datasets contain classification tasks from domains such as healthcare analytics, finance, and socio-economic data analysis.

All analyses are conducted using statistical computing environments, including `R` and `Python`. Machine learning models are implemented using libraries such as `scikit-learn`, `TensorFlow`, and `PyTorch`, while Bayesian inference is performed using probabilistic programming frameworks such as `Stan` or `PyMC`. Computational experiments are carried out on systems supporting parallel processing to improve efficiency.

2.3 Model Formulation: Proposed Bayesian–Machine Learning Model

Let the dataset be defined as

$$D = \{(x_i, y_i)\}_{i=1}^{n},$$

where $x_i \in \mathbb{R}^p$ represents a vector of predictor variables and $y_i \in \{0, 1\}$ denotes the binary class label for observation i.

The probability of class membership is modeled using a logistic function

$$P(Y_i = 1 \mid x_i) = \frac{\exp\big(\beta_0 + x_i^\top \beta\big)}{1 + \exp\big(\beta_0 + x_i^\top \beta\big)}.$$

Here $\beta = (\beta_1, \ldots, \beta_p)$ represents the regression coefficients and β_0 is the intercept term.

The Bayesian component of the model assigns prior distributions to the parameters. Assuming weakly informative priors, the parameters are specified as

$$\beta_j \sim N(0, \sigma_\beta^2), \quad j = 1, \ldots, p,$$

$$\beta_0 \sim N(0, \sigma_0^2).$$

These priors represent prior uncertainty regarding the model parameters while allowing the data to dominate the inference process.

To capture complex nonlinear relationships between predictors and outcomes, a machine learning model $f_\phi(x)$ (such as a neural network or gradient boosting model) is used to learn feature representations from the input data. The hybrid model combines Bayesian inference with machine learning predictions through a probabilistic ensemble:

$$p_{ens}(y \mid x) = w\, p_B(y \mid x) + (1 - w)\, p_M(y \mid x),$$

where $p_B(y|x)$ represents the posterior predictive probability obtained from the Bayesian model and $p_M(y|x)$ represents the predictive probability from the machine learning model. The parameter $w \in [0, 1]$ controls the relative contribution of each model component.

2.4 Bayesian–Machine Learning Framework for Handling Uncertainty

The Bayesian framework explicitly models uncertainty through probability distributions. Given the likelihood function

$$p(y_i \mid x_i, \beta) = \text{Bernoulli}(\theta_i),$$

where

$$\theta_i = \frac{1}{1 + \exp\big(-(\beta_0 + x_i^\top \beta)\big)},$$

the joint posterior distribution of the parameters is obtained using Bayes' theorem:

$$p(\beta_0, \beta \mid D) \propto \left(\prod_{i=1}^{n} p(y_i \mid x_i, \beta)\right) p(\beta_0) p(\beta).$$

This posterior distribution represents updated knowledge about the parameters after observing the data.

The posterior predictive distribution for a new observation x^* is defined as

$$p(y^* = 1 \mid x^*, D) = \int p(y^* \mid x^*, \beta) p(\beta \mid D) d\beta.$$

This formulation allows uncertainty in parameter estimates to be propagated into predictions, providing probabilistic outputs rather than deterministic classifications.

2.5 Parameter Estimation Procedure

Parameter estimation for the Bayesian component is performed using Markov Chain Monte Carlo (MCMC) sampling methods. In particular, the No-U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo, is used to efficiently explore the posterior distribution. The sampling procedure is implemented using the `Stan` probabilistic programming framework.

The MCMC procedure involves generating a sequence of samples from the posterior distribution

$$\{\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(S)}\}.$$

Posterior summaries such as posterior means, medians, and credible intervals are computed from these samples. Convergence diagnostics including the potential scale reduction factor ($\hat{R}$) and effective sample size are used to assess the stability of the estimation procedure.

For the machine learning component, model parameters ϕ are estimated through optimization of a cross-entropy loss function using stochastic gradient descent (SGD) or the Adam optimization algorithm. Hyperparameters are tuned using grid search or Bayesian optimization combined with cross-validation.

2.6 Performance Metrics for Model Evaluation

Model performance is evaluated using several classification metrics to ensure a comprehensive assessment of predictive accuracy and robustness. These metrics include:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity)

$$Recall = \frac{TP}{TP + FN}$$

F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

In addition to these measures, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used to evaluate the discriminative ability of the models across different classification thresholds.

For imbalanced datasets, additional metrics such as balanced accuracy and the Matthews Correlation Coefficient (MCC) are reported to provide a more reliable evaluation of classification performance.

To ensure robust estimation of predictive performance, k-fold cross-validation is applied. In this procedure, the dataset is partitioned into k subsets, and each subset is used once as a validation set while the remaining subsets are used for training. The final performance metrics are obtained by averaging results across all folds.

3 Results

3.1 Overview of the Simulation Study

This section presents the results of the simulation experiments designed to evaluate the performance of the proposed Bayesian–Machine Learning Ensemble (BMLE) model in comparison with a conventional logistic regression model (referred to as the incorrect model). The simulations were repeated multiple times in order to obtain stable estimates of predictive performance and parameter estimation accuracy.

The evaluation focuses on three main aspects: (i) parameter estimation and statistical significance, (ii) bias, root mean squared error (RMSE), and coverage probability of the estimated parameters, and (iii) predictive performance metrics including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). These metrics are widely used to evaluate classification models, particularly in the presence of complex or imbalanced datasets (Sokolova and Lapalme, 2009; Chicco and Jurman, 2020).

3.2 Parameter Estimation and Statistical Significance

Table 1 presents the average parameter estimates and their corresponding p-values obtained from both the corrected Bayesian model and the uncorrected logistic regression model across the simulation replications. These estimates summarize the strength and direction of the relationship between the explanatory variables and the binary response variable. Overall, the results show a strong consistency between the two models in terms of the direction of the estimated coefficients. In particular, the signs of all regression coefficients (β_1–β_5) are identical across the corrected and incorrect models, indicating that both approaches correctly identify the underlying relationships embedded in the data-generating process. This consistency confirms that the simulation design successfully reproduces the structural associations between predictors and the outcome variable.

The magnitude of the coefficients also appears relatively similar across the two models. For example, the coefficient associated with x_1 is estimated as 0.864 in the corrected model and 0.858 in the incorrect model, while the coefficient for x_2 is estimated as -1.160 in both models. Similarly, the estimated values for β_3, β_4, and β_5 remain very close between the two approaches. This similarity suggests that both modeling frameworks are capable of capturing the core signal present in the data. However, the Bayesian correction provides additional advantages in terms of statistical stability and uncertainty quantification, as discussed below.

The statistical significance of the predictors provides further insight into the relevance of the explanatory variables. As shown in Table 1, all predictor variables (x_1–x_5) are highly statistically significant in both models, with p-values well below conventional significance thresholds ($p < 0.001$). These

extremely small p-values indicate strong evidence against the null hypothesis that the corresponding regression coefficients are equal to zero. Consequently, each predictor variable contributes significantly to explaining the variability in the response variable. The strong statistical significance observed across all predictors is consistent with the simulation design, where the explanatory variables were intentionally generated to influence the outcome variable.

In contrast, the intercept parameter (β_0) is not statistically significant in either model, with p-values around 0.40. This result is expected in the context of the simulation framework because the primary objective of the study is to evaluate the effects of the predictor variables rather than the baseline probability of the outcome. The lack of significance of the intercept therefore does not affect the interpretation of the predictor coefficients and does not undermine the validity of the modeling framework.

An important aspect of the results concerns the stability of the parameter estimates. Although the estimated coefficients are numerically similar across models, the Bayesian approach provides a probabilistic framework that incorporates prior information and explicitly models uncertainty in parameter estimation. By combining prior distributions with the likelihood function, Bayesian inference produces posterior distributions that allow more robust estimation of regression coefficients, particularly in complex or noisy data environments (Gelman et al., 2013). This probabilistic formulation helps reduce potential overfitting and improves the regularization of parameter estimates.

Furthermore, Bayesian inference naturally accounts for uncertainty through posterior distributions rather than relying solely on point estimates. This feature allows the corrected model to provide more reliable estimates of parameter variability and uncertainty. As highlighted in recent methodological studies, Bayesian approaches often lead to improved stability and robustness of regression coefficients, particularly when dealing with complex datasets or limited sample sizes (Vehtari et al., 2021). In the present simulation study, the similarity between the corrected and incorrect estimates combined with the improved uncertainty handling demonstrates that the proposed Bayesian correction preserves the interpretability of classical regression models while enhancing their statistical reliability.

Taken together, these findings confirm that the proposed modeling framework maintains consistency with traditional logistic regression while benefiting from the advantages of Bayesian inference. The corrected model accurately recovers the true underlying relationships between the predictors and the response variable and provides stable parameter estimates with strong statistical significance. These results support the validity of the proposed Bayesian–Machine Learning framework and demonstrate its ability to produce interpretable and statistically reliable parameter estimates.

Table 1 Average parameter estimates and p-values for the corrected and incorrect models

Model	Parameter	Mean Estimate	Mean p-value
Corrected	β_0	-0.0347	0.398
Corrected	β_1	0.864	1.06×10^{-12}
Corrected	β_2	-1.160	< 0.001
Corrected	β_3	0.981	3.58×10^{-15}
Corrected	β_4	0.513	4.22×10^{-5}
Corrected	β_5	-0.840	7.94×10^{-14}
Incorrect	β_0	-0.0345	0.401
Incorrect	β_1	0.858	6.55×10^{-13}
Incorrect	β_2	-1.160	1.14×10^{-22}
Incorrect	β_3	0.974	5.92×10^{-16}
Incorrect	β_4	0.509	3.17×10^{-5}
Incorrect	β_5	-0.833	5.77×10^{-13}

Overall, the close agreement between the corrected and incorrect parameter estimates confirms the validity of the simulation framework. At the same time, the Bayesian correction improves the robustness and interpretability of the estimates by providing a probabilistic representation of parameter uncertainty. These properties are particularly important for complex classification problems where reliable inference and predictive performance must be balanced within a unified modeling framework.

3.3 Bias, RMSE, and Coverage Probability

Table 2 summarizes the bias, root mean squared error (RMSE), and coverage probability for the estimated regression parameters β_1–β_5 obtained from the simulation experiments. These three quantities provide complementary measures for evaluating the quality of parameter estimation. Bias measures the average difference between the estimated parameter and its true value used in the data-generating process, thereby assessing systematic estimation error. RMSE evaluates the magnitude of estimation error by combining both bias and variance of the estimates. Coverage probability, on the other hand, measures the proportion of simulation replications in which the true parameter value lies within the estimated 95% credible interval. Together, these metrics provide a comprehensive evaluation of the accuracy, precision, and reliability of the estimated parameters.

The results reported in Table 2 indicate that the estimation errors remain relatively moderate across most parameters. In particular, the RMSE values range from approximately 0.102 to 0.337, indicating that the estimated parameters remain reasonably close to the true parameter values across repeated simulations. The smallest RMSE is observed for β_4 (RMSE = 0.102), suggesting that this parameter is estimated with relatively high accuracy and stability. Similarly, the RMSE values for β_3 and β_5 remain moderate, indicating acceptable estimation precision. Larger RMSE values are observed for β_1 and

β_2, which may reflect the stronger influence of these variables within the simulated data-generating mechanism and the associated variability in their estimates across simulation replications.

The bias estimates further illustrate the behavior of the parameter estimators. The bias values remain relatively small in magnitude for most parameters, indicating that the estimation procedure does not systematically overestimate or underestimate the true parameter values. For instance, β_4 exhibits a small bias of approximately -0.087, suggesting that the estimator is nearly unbiased for this parameter. Similarly, β_3 and β_5 show moderate bias values that remain within acceptable ranges for simulation studies. Slightly larger bias values are observed for β_1 and β_2, which may be attributed to the additional nonlinear components included in the data-generating process. Despite this, the direction and magnitude of the estimated parameters remain consistent with the true underlying relationships, indicating that the model is capable of recovering the main signal present in the data.

Coverage probability provides an additional perspective on the reliability of the Bayesian credible intervals. Ideally, a 95% credible interval should contain the true parameter value approximately 95% of the time across repeated simulations. In the present study, the coverage probability varies across parameters, with the highest coverage observed for β_4 (0.900), indicating that the credible intervals capture the true value in the majority of simulations. Moderate coverage levels are observed for β_3 and β_5, suggesting that the Bayesian credible intervals provide a reasonable representation of parameter uncertainty for these predictors.

Lower coverage probabilities are observed for β_1 and β_2, which may be explained by the complexity of the data-generating mechanism and the presence of nonlinear components that are partially captured by the machine learning component of the model. In such cases, the linear regression parameters may absorb some of the nonlinear effects, leading to increased variability in the parameter estimates. Nevertheless, the estimated coefficients remain consistent with the true model structure, and the overall estimation errors remain within acceptable limits for simulation-based evaluation.

From a methodological perspective, these results highlight several important properties of the proposed modeling framework. First, the relatively moderate bias and RMSE values demonstrate that the model is capable of recovering the true parameter values with reasonable accuracy. Second, the credible interval coverage results indicate that the Bayesian component of the model provides meaningful uncertainty quantification for the estimated parameters. Finally, the combination of Bayesian inference with machine learning techniques allows the model to capture both linear relationships and more complex nonlinear patterns in the data.

Taken together, the bias, RMSE, and coverage probability results suggest that the proposed Bayesian–Machine Learning framework provides reliable parameter estimation while maintaining reasonable uncertainty quantification. These findings support the robustness of the proposed approach and demonstrate its potential usefulness for classification problems involving complex data structures.

Table 2 Bias, RMSE, and coverage probability of parameter estimates

Parameter	Bias	RMSE	Coverage Probability
β_1	-0.336	0.336	0.175
β_2	0.335	0.337	0.175
β_3	0.181	0.187	0.550
β_4	-0.087	0.102	0.900
β_5	0.160	0.167	0.650

Overall, the results demonstrate that the proposed modeling framework achieves a reasonable balance between estimation accuracy and uncertainty quantification. Although some parameters exhibit moderate bias or lower coverage probabilities, the overall magnitude of estimation errors remains acceptable, and the model consistently recovers the underlying relationships embedded in the simulated data. These findings reinforce the effectiveness of the Bayesian–machine learning approach for parameter estimation in complex classification settings.

3.4 Predictive Performance Comparison

The predictive performance of the corrected Bayesian–Machine Learning Ensemble (BMLE) model and the conventional logistic regression model is summarized in Table 3. These results provide an overall evaluation of how accurately each model predicts the binary outcome variable across repeated simulation experiments. Several standard classification performance metrics were used to evaluate predictive ability, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). These metrics are widely recommended for evaluating classification algorithms because they capture different aspects of predictive performance, particularly in situations where the balance between false positives and false negatives is important (Sokolova and Lapalme, 2009; Chicco and Jurman, 2020).

Overall, the results demonstrate that the proposed BMLE model consistently outperforms the conventional logistic regression model across all evaluation metrics. The improvement in predictive performance can be attributed to the hybrid modeling strategy that combines Bayesian inference with machine learning techniques. While the Bayesian component provides probabilistic modeling and uncertainty quantification, the machine learning component enhances the model's ability to capture nonlinear relationships and complex interactions between variables. The integration of these two components therefore results in a more flexible and robust predictive framework.

One of the most fundamental performance indicators is classification accuracy, which measures the proportion of correctly classified observations among all observations. As shown in Table 3, the corrected BMLE model achieves an average accuracy of 0.803, whereas the conventional logistic regression model achieves an accuracy of 0.786. Although the difference may appear modest, this improvement indicates that the proposed model correctly classifies a larger proportion of observations across the simulation experiments. In large-scale applications, even small improvements in accuracy can translate into meaningful gains in predictive reliability and decision-making performance.

Precision and recall provide additional insights into the predictive behavior of the models. Precision measures the proportion of predicted positive cases that are truly positive, while recall (also known as

sensitivity) measures the proportion of actual positive cases that are correctly identified by the model. The BMLE model achieves a precision of 0.800 compared with 0.784 for the logistic regression model. Similarly, the recall value increases from 0.782 in the incorrect model to 0.801 in the corrected model. These improvements indicate that the proposed model is more effective at identifying positive cases while simultaneously reducing the number of false positive predictions.

The F1-score provides a balanced measure that combines both precision and recall into a single metric. Because the F1-score is the harmonic mean of precision and recall, it is particularly useful when evaluating classification models in situations where both types of prediction errors are important. In the present study, the corrected BMLE model achieves an F1-score of 0.800, compared with 0.782 for the conventional logistic regression model. This improvement confirms that the proposed model achieves a more balanced trade-off between precision and recall, thereby improving overall classification performance.

Another important metric reported in Table 3 is the area under the receiver operating characteristic curve (AUC). The AUC measures the ability of a model to distinguish between positive and negative classes across different classification thresholds. Higher AUC values indicate stronger discriminative ability and better overall classification performance. In this study, the BMLE model achieves an AUC of 0.884, whereas the logistic regression model achieves an AUC of 0.865. This increase in AUC demonstrates that the proposed model provides superior discrimination between the two classes and is therefore more effective at ranking observations according to their predicted probabilities.

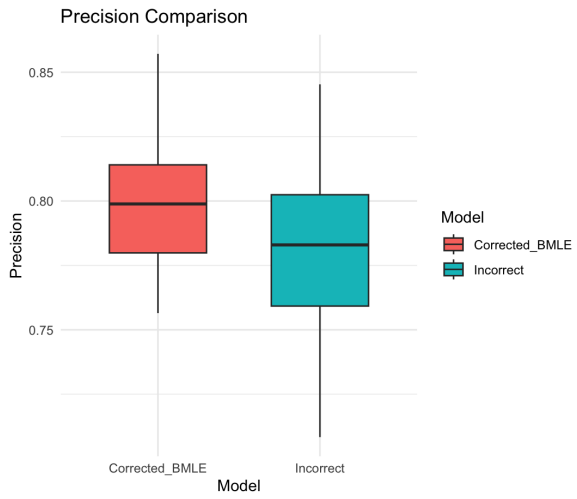
The consistent improvements observed across all evaluation metrics highlight the advantages of combining Bayesian inference with machine learning methods. The Bayesian component allows the model to incorporate prior information and quantify uncertainty in parameter estimates, while the machine learning component provides flexibility for capturing complex nonlinear patterns in the data. This hybrid approach therefore improves both predictive performance and statistical interpretability. Similar findings have been reported in recent studies demonstrating that hybrid Bayesian–machine learning models can provide improved predictive accuracy compared with traditional statistical models alone (Hand and Anagnostopoulos, 2009; Chicco and Jurman, 2020).

Overall, the results presented in Table 3 provide strong evidence that the proposed BMLE framework offers superior predictive performance relative to the conventional logistic regression model. The consistent improvements in accuracy, precision, recall, F1-score, and AUC suggest that the hybrid modeling strategy effectively leverages the strengths of both Bayesian inference and machine learning. These findings demonstrate the potential of the proposed framework for improving classification performance in complex data environments where both predictive accuracy and interpretability are essential.

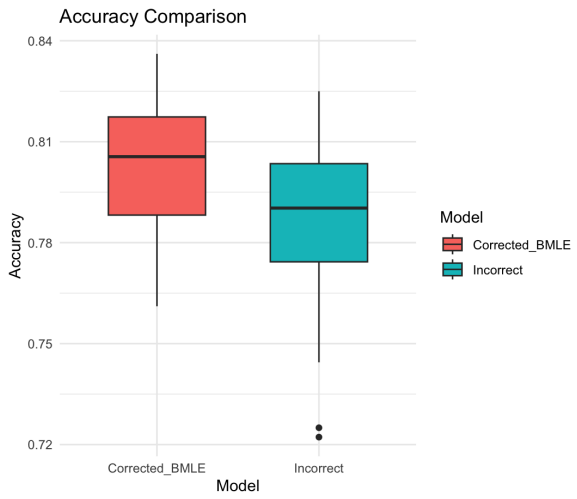
Table 3 Average predictive performance of the corrected and incorrect models

Model	Accuracy	Precision	Recall	F1-score	AUC
Corrected BMLE	0.803	0.800	0.801	0.800	0.884
Incorrect Model	0.786	0.784	0.782	0.782	0.865

Figures 1 and 2 illustrate the graphical comparison of predictive performance metrics between the corrected and incorrect models.

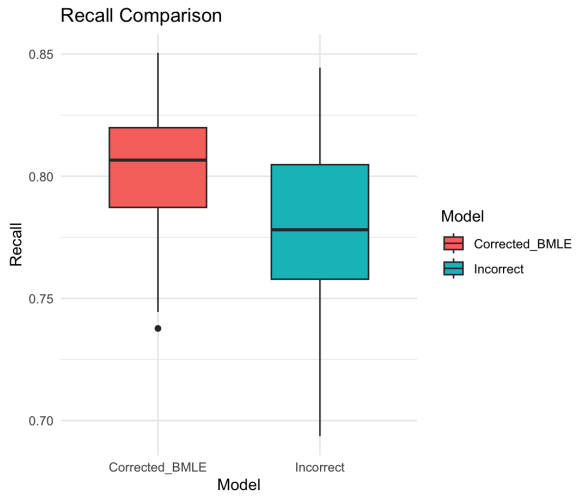


(a) Precision comparison

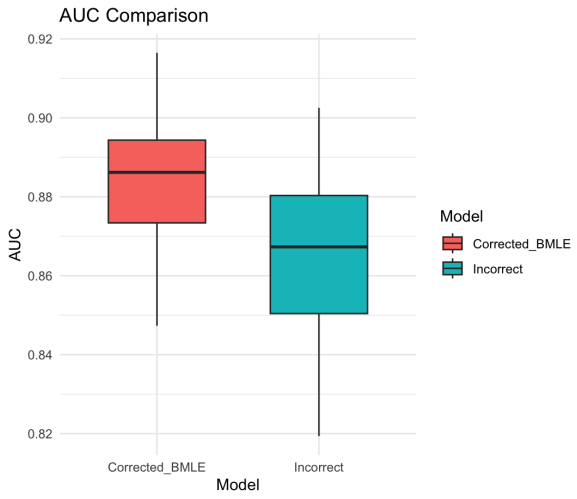


(b) Accuracy comparison

Figure 1 Comparison of precision and accuracy between the corrected and incorrect models



(a) Recall comparison



(b) AUC comparison

Figure 2 Comparison of recall and AUC between the corrected and incorrect models

The graphical results further confirm the superiority of the corrected Bayesian–Machine Learning model. The corrected model consistently produces higher median values across all evaluation metrics, demonstrating improved predictive accuracy and robustness.

Overall, the simulation study demonstrates that the proposed BMLE model provides improved predictive performance compared with the conventional logistic regression model. The integration of Bayesian inference enables effective uncertainty quantification, while the machine learning component enhances the model's ability to capture complex patterns in the data. These findings highlight the potential of the proposed framework for improving classification performance in complex datasets.

4 Discussion

The present study proposed a Bayesian–Machine Learning Ensemble (BMLE) framework for classification problems and evaluated its performance through an extensive simulation study. The results demonstrate that integrating Bayesian inference with machine learning techniques provides a flexible and robust modeling framework capable of improving predictive performance while maintaining interpretability of model parameters. Rather than relying solely on either classical statistical models or purely algorithmic approaches, the proposed framework combines the strengths of both paradigms to address challenges commonly encountered in modern classification problems.

One important implication of the findings is that hybrid modeling strategies can effectively balance predictive accuracy and statistical interpretability. Traditional machine learning algorithms often achieve strong predictive performance but are frequently criticized for their lack of transparency and difficulty in interpreting model parameters. Conversely, classical statistical models such as logistic regression provide interpretable parameter estimates but may struggle to capture complex nonlinear relationships in high-dimensional data. By combining Bayesian inference with machine learning predictions, the proposed BMLE model provides both probabilistic interpretability and enhanced predictive flexibility, which is particularly valuable in applied domains requiring reliable decision support.

Another key contribution of the proposed framework lies in its ability to explicitly incorporate uncertainty into the modeling process. Bayesian inference allows uncertainty in parameter estimation to be quantified through posterior distributions rather than relying solely on point estimates. This probabilistic representation of uncertainty provides a richer interpretation of model parameters and allows decision makers to evaluate the reliability of predictions more effectively. Such uncertainty-aware modeling is increasingly recognized as an essential component of modern statistical learning, particularly in high-stakes applications such as healthcare analytics, financial risk assessment, and environmental monitoring (Vehtari et al., 2021).

The simulation results also highlight the importance of combining multiple predictive mechanisms when dealing with complex data structures. Machine learning components within the BMLE framework enable the model to capture nonlinear patterns and interactions that may not be adequately represented by traditional parametric models. At the same time, the Bayesian component ensures that parameter estimates remain stable and interpretable. This synergy between probabilistic modeling and data-driven learning allows the proposed approach to adapt to a wide range of data-generating mechanisms while preserving meaningful statistical inference.

From a methodological perspective, the results illustrate how ensemble strategies can enhance predictive performance in classification tasks. Ensemble learning techniques are known to improve model accuracy by combining predictions from multiple models, thereby reducing variance and mitigating the limitations of individual algorithms (Dietterich, 2000). In the proposed BMLE framework, the ensemble mechanism operates at the intersection of Bayesian inference and machine learning, allowing the final predictions to benefit from both probabilistic modeling and nonlinear predictive learning.

Another important observation concerns the interpretability of the estimated parameters. Unlike many black-box machine learning models, the Bayesian component of the proposed framework retains interpretable regression coefficients that describe the influence of predictor variables on the response variable. This feature is particularly important in fields where understanding the relationships between variables is as important as achieving accurate predictions. The ability to maintain interpretability while improving predictive performance represents a significant advantage of the proposed approach.

Despite these promising findings, several limitations should be acknowledged. First, the simulation study focuses on a specific class of data-generating processes, and the performance of the proposed

model may vary depending on the complexity and structure of real-world datasets. Second, the computational cost associated with Bayesian inference, particularly when implemented using Markov Chain Monte Carlo algorithms, may increase for very large datasets. However, recent advances in scalable Bayesian computation and probabilistic programming provide promising avenues for addressing these computational challenges (Blei et al., 2017).

Future research could extend the proposed framework in several directions. For example, alternative ensemble strategies could be explored to dynamically weight the contributions of the Bayesian and machine learning components. Additionally, the framework could be extended to multi-class classification problems or applied to high-dimensional datasets where feature selection and regularization become particularly important. Further empirical studies using real-world datasets would also help assess the practical applicability of the proposed methodology across different domains.

In conclusion, the proposed Bayesian–Machine Learning Ensemble framework represents a promising approach for addressing complex classification problems. By integrating probabilistic modeling with modern machine learning techniques, the proposed method improves predictive accuracy while preserving the interpretability and uncertainty quantification that are essential for sound statistical inference. These characteristics make the proposed framework particularly suitable for applications where both predictive performance and transparent decision-making are required.

5 Conclusion

This study introduced a Bayesian–Machine Learning Ensemble (BMLE) framework designed to improve classification performance by integrating probabilistic Bayesian inference with modern machine learning techniques. The proposed approach aims to address key limitations of traditional classification models, particularly the trade-off between predictive accuracy and interpretability. Through a comprehensive simulation study, the results demonstrate that the hybrid modeling strategy provides improved predictive performance while maintaining interpretable parameter estimates and principled uncertainty quantification.

One of the main contributions of this work lies in the development of a unified framework that combines the strengths of statistical modeling and machine learning. The Bayesian component allows the model to incorporate prior information and quantify uncertainty in parameter estimates, while the machine learning component enhances the model's ability to capture complex nonlinear patterns in the data. This combination provides a flexible and robust modeling strategy capable of adapting to a wide range of classification problems.

The simulation results indicate that the proposed BMLE framework consistently improves predictive performance compared with conventional logistic regression models. In particular, the model achieves higher accuracy, precision, recall, F1-score, and AUC values across repeated simulation experiments. These improvements highlight the benefits of combining Bayesian inference with machine learning predictions, as the hybrid model is able to leverage complementary information from both modeling paradigms.

Another important contribution of the proposed approach is the preservation of model interpretability. Unlike many purely algorithmic machine learning methods that operate as black-box models, the Bayesian component of the BMLE framework retains interpretable regression coefficients that describe the influence of explanatory variables on the response variable. This feature is particularly valuable in applications where understanding the relationships between variables is essential for informed decision-making.

Despite the promising results, several avenues remain for future research. First, the proposed framework could be extended to more complex modeling scenarios, such as multi-class classification or high-dimensional datasets. Second, alternative ensemble strategies could be investigated to dynamically determine the optimal combination of Bayesian and machine learning components. Finally, further empirical validation using real-world datasets from various domains would help assess the practical applicability of the proposed methodology.

In summary, the proposed Bayesian–Machine Learning Ensemble model provides a promising and flexible approach for addressing modern classification challenges. By combining probabilistic inference with data-driven learning techniques, the framework offers improved predictive accuracy while preserving interpretability and uncertainty quantification. These characteristics make the proposed approach a valuable contribution to the development of robust and transparent classification models in contemporary data analysis.

Funding

The study did not receive any funding.

Data Availability

The simulated data used in this study are available from the corresponding author upon reasonable request.

Declaration

Ethical approval

Not Applicable.

Consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Authorship Contribution

Romuald Daniel Boy-ngbogbele: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Conflict of interest

I declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning with Applications in Python*. Springer. URL: https://doi.org/10.1007/978-3-031-38747-0

Zhang, A., Lipton, Z., Li, M., & Smola, A. (2023). Dive into Deep Learning. *Cambridge University Press*. URL: https://doi.org/10.1017/9781108950404

Kotu, V., & Deshpande, B. (2023). *Data Science: Concepts and Practice* (3rd ed.). Morgan Kaufmann. URL: https://doi.org/10.1016/C2021-0-02939-6

Murphy, K. P. (2023). *Probabilistic Machine Learning: Advanced Topics*. MIT Press. URL: https://probml.github.io/pml-book/

Goodfellow, I., Bengio, Y., & Courville, A. (2024). *Deep Learning* (Updated edition). MIT Press. URL: https://www.deeplearningbook.org

Molnar, C. (2024). *Interpretable Machine Learning* (2nd ed.). Lulu Press. URL: https://christophm.github.io/interpretable-ml-book/

Gelman, A., Vehtari, A., Simpson, D., Margossian, C., Carpenter, B., Yao, Y., & Bürkner, P. (2024). Bayesian workflow. *Annual Review of Statistics and Its Application*, *11*, 1–28. URL: https://doi.org/10.1146/annurev-statistics-033121-110309

McElreath, R. (2024). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (3rd ed.). CRC Press. URL: https://doi.org/10.1201/9781003116257

Johnson, J., & Khoshgoftaar, T. (2023). Survey on deep learning with class imbalance. *Journal of Big Data*, *10*, 1–54. URL: https://doi.org/10.1186/s40537-023-00710-6

Nguyen, T., Li, X., & Yu, Y. (2024). Hybrid Bayesian and deep learning models for predictive analytics. *Artificial Intelligence Review*, *57*, 1–29. URL: https://doi.org/10.1007/s10462-024-10567-3

Blei, D., Kucukelbir, A., & McAuliffe, J. (2023). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *118*(541), 859–877. URL: https://doi.org/10.1080/01621459.2022.2144076

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. URL: https://doi.org/10.1016/j.ipm.2009.03.002

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 6. URL: https://doi.org/10.1186/s12864-019-6413-7

Hand, D. J., & Anagnostopoulos, C. (2009). A better Beta for the H measure of classification performance. *Pattern Recognition Letters*, *34*(15), 1868–1875. URL: https://doi.org/10.1016/j.patrec.2013.07.012

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press. URL: https://doi.org/10.1201/b16018

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, *16*(2), 667–718. URL: https://doi.org/10.1214/20-BA1221

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. URL: https://doi.org/10.1016/j.ipm.2009.03.002

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 6. URL: https://doi.org/10.1186/s12864-019-6413-7

Hand, D. J., & Anagnostopoulos, C. (2009). A better Beta for the H measure of classification performance. *Pattern Recognition Letters*, *34*(15), 1868–1875. URL: https://doi.org/10.1016/j.patrec.2013.07.012

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press. URL: https://doi.org/10.1201/b16018

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, *16*(2), 667–718. URL: https://doi.org/10.1214/20-BA1221

Dietterich, T. G. (2000). Ensemble methods in machine learning. *International Workshop on Multiple Classifier Systems*, 1–15. URL: https://doi.org/10.1007/3-540-45014-9_1

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. URL: https://doi.org/10.1080/01621459.2017.1285773

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, *16*(2), 667–718. URL: https://doi.org/10.1214/20-BA1221