Communication in Physical Sciences, 2026, 13(8):1385-1393

A Comparative Analysis of Sensitivity and Specificity of LASSO Logistic Regression and Random Forest Using Sonar Data

Fortune Meka.
Received: 21 March 2026/Accepted: 20 July 2026/Published: 28 July 2026
https://doi.org/10.4314/cps.v13i8.10

Abstract: *In this study we carried out a comparison of the sensitivity and specificity of LASSO Logistic Regression and Random Forest in the binary classification of Sonar observations. Here, we utilized the Sonar dataset available in the mlbench R package, comprising 208 observations and 60 predictor variables, with observations classified into Mine and Rock categories. The data were split into training and testing sets using a stratified 80:20 split. LASSO Logistic Regression was fitted using 10-fold cross-validation to determine the optimal regularization parameter, while Random Forest was implemented using 500 decision trees. The predictive performance of the models was evaluated using confusion matrices, with sensitivity and specificity serving as the primary performance measures. McNemar's test was employed to assess whether the two models produced significantly different classification decisions on the same test observations. The results revealed that Random Forest demonstrated higher sensitivity, while both models achieved identical specificity. McNemar's test revealed that the difference in the paired classification decisions was not statistically significant at the 5% level.*

Keywords: *LASSO Logistic Regression; Random Forest; Sensitivity; Specificity; Sonar Data; McNemar's Test.*

Fortune Meka
Department of Mathematics and Statistics, University of Delta, Agbor.
Email: fortune.meka@unidel.edu.ng

1.0 Introduction

Binary classification is a fundamental task in statistical learning and machine learning, with applications in medical diagnosis, financial risk assessment, signal detection, pattern recognition, and other domains where observations must be assigned to one of two outcome classes.. Such classification problems arise in a broad range of applications, such as medical diagnosis, financial risk assessment, pattern recognition and many more. The quality of a binary classification model cannot, however, be correctly determined by simply examining whether observations are correctly or incorrectly classified. Different types of classification errors may have different implications, making it necessary to examine measures that distinguish between correctly identified positive and negative observations. Among the most important measures are sensitivity and specificity, which provide complementary information about the ability of a classification model to correctly identify observations belonging to the two outcome classes (Altman & Bland, 1994). Consequently, sensitivity and specificity provide complementary measures for evaluating binary classifiers because they distinguish the ability of a model to correctly identify positive observations from its ability to correctly identify negative observations.Sensitivity reflects the ability of a classifier to detect positive cases, while specificity reflects its ability to correctly exclude negative cases (Altman & Bland, 1994). These measures are derived from the confusion matrix, which summarizes true-positive, true-negative, false-positive, and

false-negative classifications. A model may achieve relatively high overall accuracy while performing poorly in identifying one of the two classes. The separate consideration of sensitivity and specificity is what gives clear informative assessment of classification behavior, particularly when the consequences of false-positive and false-negative classifications differ.

Logistic Regression is a widely used statistical classification method in which the probability of a binary outcome is modelled as a function of one or more explanatory variables through the logistic transformation.. This approach is quite valuable because it provides an interpretable relationship between predictors and the probability of membership of a group. The estimated coefficients can be transformed into odds ratios, this allows one to examine the direction and magnitude of associations between explanatory variables and the binary response (Hosmer *et al*., 2013). Logistic Regression is also an important component of the statistical-learning framework and is commonly used as a benchmark classification method because of its build-up is simple and easy to interpret (Hastie *et al*., 2009).

When the underlying data structure is more complex than this specification permits, the resulting classification performance may be affected. However, conventional Logistic Regression may become less effective when the number of predictors is relatively large, when predictors are strongly correlated, or when several variables contribute little information to classification. LASSO (Least Absolute Shrinkage and Selection Operator) Logistic Regression addresses these challenges by incorporating an L1 regularization penalty into the model estimation process. The penalty shrinks regression coefficients toward zero and can reduce some coefficients exactly to zero, thereby performing variable selection while controlling model complexity (Tibshirani, 1996). LASSO Logistic Regression is therefore particularly useful in classification problems involving numerous predictors relative to the available observations. Modern statistical learning has therefore introduced more flexible methods that can learn complex relationships from the data without requiring the researcher to specify all possible nonlinearities and interactions in advance (Hastie *et al*., 2009). This property is particularly relevant to the Sonar dataset, which contains 208 observations and 60 predictor variables. The relatively large number of predictors compared with the sample size provides a suitable setting for examining whether regularization and variable selection improve or alter classification performance relative to a more flexible ensemble-learning approach such as Random Forest.

An alternative approach that does not impose the same linear modelling assumptions is Random Forest (Breiman, 2001).

Random Forest constructs a large collection of decision trees using random samples of observations and randomly selected subsets of predictor variables, and combines the predictions from the individual trees to obtain a final classification. The approach can model nonlinear relationships and interactions among predictors without requiring them to be explicitly specified. Breiman (2001) demonstrated that Random Forest can produce strong generalization performance. The contrasting characteristics of Logistic Regression and Random Forest have generated considerable interest in determining whether the greater flexibility of machine-learning methods results in better classification performance. Couronné *et al*., (2018) conducted a large-scale experiment using 243 real datasets to compare Random Forest and Logistic Regression as binary classification methods. Their results showed that Random Forest achieved higher accuracy than Logistic Regression on a substantial proportion of the datasets and generally performed favorably according to several predictive performance measures. However, the authors did not

conclude that Random Forest was universally superior. Rather, the relative performance of the two methods varied across datasets and depended on characteristics of the data and modelling problem. These findings emphasize that classifier performance is dataset-dependent and that no single modelling approach can be assumed to be universally superior..

Comparative evidence also indicates that the relative performance of statistical and machine-learning models depends on the characteristics of the data and the criteria used for evaluation. Christodoulou et al. (2019) reported that Logistic Regression often performed competitively, particularly when analyses were restricted to studies with a lower risk of bias, suggesting that increased model complexity does not necessarily guarantee superior predictive performance. These findings reinforce the importance of empirical, dataset-specific comparisons rather than assuming that flexible machine-learning algorithms will consistently outperform conventional statistical models. Comparative evidence also indicates that the relative performance of statistical and machine-learning models depends on the characteristics of the data and the criteria used for evaluation. Christodoulou et al. (2019) reported that Logistic Regression often performed competitively, particularly when analyses were restricted to studies with a lower risk of bias, suggesting that increased model complexity does not necessarily guarantee superior predictive performance. These findings reinforce the importance of empirical, dataset-specific comparisons rather than assuming that flexible machine-learning algorithms will consistently outperform conventional statistical models. Regression when studies with a low risk of bias were considered. Other researchers also supported empirical comparisons (Altman & Bland, 1994; Fawcett, 2006). According to Garett *et al*., (2013), it is important to understand the difference between different statistical approaches. The two approaches differ fundamentally in their modelling assumptions, complexity, interpretability, and ability to represent nonlinear relationships. Thus, LASSO Logistic Regression and Random Forest represent contrasting modelling strategies. LASSO Logistic Regression retains the interpretability of a regression-based classifier while controlling model complexity through regularization, whereas Random Forest uses an ensemble of decision trees to capture nonlinear relationships and interactions among predictors. Comparing these approaches provides an opportunity to determine whether greater model flexibility translates into superior classification of the Sonar observations.

The present study uses the Sonar dataset, a benchmark binary classification dataset available through the mlbench R package. The dataset contains 208 observations and 60 signal-related predictor variables, with observations classified into Mine and Rock categories. Its relatively high predictor-to-observation ratio makes it suitable for examining the performance of a regularized Logistic Regression model alongside a flexible ensemble-learning method. LASSO Logistic Regression provides an interpretable and regularized statistical benchmark, whereas Random Forest provides a nonlinear modelling alternative capable of capturing complex interactions among predictors. Logistic Regression provides an interpretable statistical benchmark, while Random Forest provides a flexible alternative capable of accommodating nonlinear patterns and interactions.

Although previous research has extensively compared Logistic Regression and Random Forest, much of the existing literature has emphasized overall predictive performance or broad benchmarking across multiple datasets. Couronné *et al*. (2018), for example, focused on large-scale comparisons of predictive performance across 243 datasets, while

Christodoulou *et al*. (2019) examined the broader question of whether machine-learning methods provide a general performance advantage over Logistic Regression. There is therefore the need of a focused empirical investigation that examines the two methods specifically through the complementary measures of sensitivity and specificity using a recognized binary classification dataset. Such an analysis can reveal whether one model is better at identifying observations belonging to one class while the other may perform better in correctly identifying the alternative class.

The aim of this study is to comparatively evaluate the sensitivity and specificity of LASSO Logistic Regression and Random Forest for binary classification using the Sonar dataset.

.

Specifically, the study seeks to: (i) develop a LASSO Logistic Regression model using the Sonar dataset; (ii) develop a Random Forest classifier using the same training and testing observations; (iii) determine the sensitivity and specificity of each model; and (iv) compare their paired classification decisions using McNemar's test.

The study determines the sensitivity and specificity of each model, and compares the two measures across the competing classification approaches. The study is significant because it provides a focused assessment of two fundamentally different classification strategies using sensitivity and specificity rather than relying solely on overall accuracy. The findings will contribute empirical evidence on whether the flexibility of Random Forest produces different sensitivity and specificity characteristics from LASSO Logistic Regression and may assist researchers in selecting appropriate classification methods when false-positive and false-negative errors have different implications..

More broadly, the study demonstrates the importance of evaluating classification algorithms using multiple performance measures and selecting models according to the characteristics of the dataset and the specific consequences associated with classification errors.

2. 0 Methodology

2.1 LASSO Logistic Regression

Logistic Regression is appropriate when the dependent variable is binary. Let Y_i denote the binary response for observation i, where $Y_i = 1$ represents the Mine (M) class and $Y_i = 0$ represents the Rock ® class. For p predictor variables $X_i, \dots, X_p$, the probability that observation i belongs to the positive class was modelled using equation 1

$$P(Y_i = 1|X_i) = \frac{exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in})}{1 + exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in})} \quad (1)$$

Because the Sonar dataset contains 60 predictor variables relative to 208 observations, LASSO regularization was used to control model complexity and improve estimation stability. LASSO (Least Absolute Shrinkage and Selection Operator) introduces an L_1 penalty that shrinks some regression coefficients toward zero and can set some coefficients exactly equal to zero, thereby performing variable selection (Tibshirani, 1996).

For binary Logistic Regression, the LASSO estimator can be expressed according to equation 2

$$\hat{\beta} = \underset{\beta}{\text{Arg min}} \left\{ -\ell(\beta) + \lambda \sum_{j=1}^{n} |\beta_j| \right\} \quad (2)$$

where $\ell(\beta)$ is the logistic log-likelihood, $\lambda \geq 0$ is the regularization parameter, and $\sum_{j=1}^{n} |\beta_j|$ is the L_1 penalty.

The value of λ controls the degree of shrinkage. When $\lambda = 0$, there is no LASSO penalty, whereas increasing λ produces greater shrinkage of the coefficients. In this study, the

optimal value of λ was chosen through 10-fold cross-validation using the training data. The value of λ that minimized the cross-validation classification error ($\boldsymbol{\lambda}$ min) was used for prediction. The LASSO formulation follows the original proposal of Tibshirani (1996).

For classification, the estimated probability was converted into a predicted class using a 0.50 probability threshold expressed in equation 3

$$\hat{Y}_i = \begin{cases} 1, & P(Y_i = 1|X_i) \geq 0.5 \\ 0, & P(Y_i = 1|X_i) < 0.5 \end{cases} \quad (3)$$

Thus, observations with predicted probability of at least 0.50 were classified as Mine, while those below 0.50 were classified as Rock.

2.2 Random Forest

Random Forest is an ensemble classification method introduced by Breiman (2001). It constructs a large number B of decision trees using bootstrap samples of the training observations and randomly selected subsets of predictors at each node. The final classification is obtained by aggregating the predictions of the individual trees through majority voting (Breiman, 2001).

Let $T_b(X), b = 1,2, \dots, b$ denote the classification produced by the b-th decision tree for an observation with predictor vector X. The Random Forest classification was obtained as

$$\hat{C}(X) = Arg \max_{c \in \{M,R\}} \sum_{b=1}^{B} I(T_b(X) = c) \quad (4)$$

where $I(\cdot)$ is an indicator function that takes the value 1 when the condition is true and 0 otherwise.

In this study, the Random Forest model was fitted using 500 trees. At each split, a randomly selected subset of predictor variables was considered for determining the split. The final class assignment was based on the majority vote across the 500 trees. This follows the fundamental Random Forest framework proposed by Breiman (2001), in which randomization of observations and predictor variables is combined with aggregation across many trees.

2.3 Evaluation of Predictive Performance

The classification results from both models were summarised using a 2×2 confusion matrix.

Table 1: Model Confusion Matrix

	Actual Positive Mine	**Actual Negative Rock**
Predicted Positive Mine	True Positive(TP)	False Positive(FP)
Predicted Negative Rock	False Negative(FN)	True Negative(TN)

$$Sensitivity = \frac{TP}{TP + FN} \quad (5)$$

A high sensitivity indicates that the model is effective at detecting positive observations and produces relatively few false-negative classifications Altman and Bland (1994).

$$Specitivity = \frac{TN}{TN + FP} \quad (6)$$

A high specificity indicates that the model is effective at correctly identifying negative observations and produces relatively few false-positive classifications (Altman & Bland, 1994)

It is safe to conclude that the model with the higher sensitivity will be considered superior in identifying positive observations, whereas the model with the higher specificity will be considered superior in identifying negative observations.

Because Logistic Regression and Random Forest were applied to the same test observations, their classification results are paired rather than independent. The McNemar's test proposed by McNemar (1947) is therefore appropriate for assessing whether the two classifiers produce significantly

different classification decisions on paired observations.

The paired classification results can be arranged as

Table 2: Model Paired Classification Result

	Random Forest Correct	*Random Forest Incorrect*
LASSO Correct	p	Q
LASSO Incorrect	r	S

The McNemar chi-square statistic with continuity correction is

$$\chi^2 = \frac{(|q-r|-1)^2}{q+r} \quad (7)$$

A p-value below 0.05 points at a statistically significant evidence that the two models produce different paired classification decisions, whereas a p-value of 0.05 or greater reveals insufficient evidence of a difference.

2.4 Description of Dataset

The Sonar dataset (Gorman & Sejnowski, 1988), a benchmark dataset which investigated the classification of sonar returns from two types of underwater objects: a metal cylinder and a similarly shaped rock. The dataset contains sonar signal measurements used to distinguish between the two classes. It contains 208 observations and 60 continuous predictor variables, creating a classification problem in which the number of predictors is relatively large compared with the number of observations. All analysis was carried out using the R software package.

3. 0 Results and Discussions

3.1 Description of Dataset

The Sonar dataset comprised 208 observations and 61 variables, including 60 signal-related predictor variables and one binary response variable (Class). The response variable consisted of two classes, Mine (M) and Rock (R). Of the 208 observations, 111 (53.37%) were classified as Mine and 97 (46.63%) as Rock. As shown in Table 3, the two classes were reasonably balanced, with Mine observations accounting for 53.37% and Rock observations accounting for 46.63% of the dataset.

Table 3. Class distribution of the Sonar dataset

Class	*Frequency*
Mine	111
Rock	97
Total	208

3.2 Sensitivity and Specificity of LASSO Logistic Regression

As presented in Table 4, LASSO Logistic Regression correctly classified 15 of the 22 Mine observations and 15 of the 19 Rock observations in the test set. The model therefore achieved a sensitivity of 68.18% and a specificity of 78.95%. Thus, the model correctly identified 68.18% of the actual Mine observations The resulting overall classification accuracy was 73.17%, indicating that 30 of the 41 test observations were correctly classified.The confusion matrix presented in Table 4 shows that 15 Mine observations were correctly classified as Mine, while 7 Mine observations were incorrectly classified as Rock. Also, 15 Rock observations were correctly classified as Rock, while 4 Rock observations were incorrectly classified as Mine.

Table 4: Confusion Matrix of LASSO Logistic Regression

	Actual Mine	*Actual Rock*
Predicted Mine	15	4
Predicted Rock	7	15

3.3 Sensitivity and Specificity of Random Forest

As shown in Table 5, Random Forest correctly classified 20 of the 22 Mine observations and

15 of the 19 Rock observations. The model therefore achieved a sensitivity of 90.91% and a specificity of 78.95%. This indicates that Random Forest correctly identified approximately 91% of the actual Mine observations while correctly identifying approximately 79% of the actual Rock observations. The Random Forest model achieved an overall accuracy of 85.37%, corresponding to 35 correctly classified observations out of the 41 test observations. This was higher than the 73.17% accuracy obtained by LASSO Logistic Regression.

Table 5: Confusion Matrix of Random Forest

	Actual Mine	*Actual Rock*
Predicted Mine	20	4
Predicted Rock	2	15

Sensitivity was 90.91% while specificity was 78.95% for the Random forest**.** Therefore, the model correctly identified approximately 91% of the actual Mine observations, while correctly identifying approximately 79% of the actual Rock observations.

The Random Forest model also achieved an overall accuracy of 85.37%**,** which was higher than the 73.17% obtained by LASSO Logistic Regression.

3.4 Comparative Analysis of Sensitivity and Specificity

As shown in Table 6, Random Forest achieved substantially higher sensitivity than LASSO Logistic Regression (90.91% versus 68.18%), representing an absolute difference of 22.73 percentage points. However, both models achieved identical specificity of 78.95%.

Table 5: Comparison of Sensitivity and Specificity

	Sensitivity(%)	Specificity (%)
LASSO Logistic Regression	68.18	78.95
Random Forest	90.91	78.95

These findings indicate that the principal difference between the models was their ability to identify Mine observations, whereas their ability to correctly classify Rock observations was identical in this test set.The higher sensitivity of Random Forest may reflect its ability to capture nonlinear relationships and interactions among the 60 signal-related predictors that may not be adequately represented by the regularized linear structure of LASSO Logistic Regression. However, this interpretation should be treated cautiously because the comparison is based on a relatively small test set of 41 observations.

3.5. Comparison of Classification Decisions

McNemar's test was used to compare the paired classification decisions produced by the two models on the same test observations.

As shown in Table 7, the McNemar test produced a chi-square statistic of 2.286 with 1 degree of freedom and a p-value of 0.131.

Table7: McNemar's Test Results

Statistic	*McNemar's* χ^2	*df*	*p-value*
Value	2.286	1	0.131

Because the p-value (0.131) exceeds the prespecified significance level of 0.05, the null hypothesis of no difference in the paired classification decisions is not rejected. Thus, the analysis did not detect a statistically significant difference between the classification decisions of LASSO Logistic Regression and Random Forest on the test observations at the 5% significance level.Although Random Forest showed a markedly higher observed sensitivity than LASSO Logistic Regression, the McNemar test did not detect a statistically significant difference in the paired classification decisions. The apparent difference in sensitivity should

therefore be interpreted as an observed performance difference in this test sample rather than evidence of a statistically significant superiority of Random Forest. The relatively small test-set size may also limit the statistical power to detect differences between the models.

4.0 Conclusion

This study comparatively evaluated the sensitivity and specificity of LASSO Logistic Regression and Random Forest for binary classification using the Sonar dataset. Random Forest achieved a higher sensitivity of 90.91%, compared with 68.18% for LASSO Logistic Regression, while both models recorded an identical specificity of 78.95%. Random Forest also achieved higher overall accuracy (85.37%) than LASSO Logistic Regression (73.17%). However, McNemar's test indicated that the difference in the paired classification decisions was not statistically significant at the 5% level ($\chi^2 = 2.286$, $p = 0.131$).

These findings suggest that Random Forest provided better observed performance in identifying Mine observations in the test sample, whereas both models performed similarly in identifying Rock observations. Therefore, Random Forest may be considered when sensitivity is a primary performance requirement. Nevertheless, LASSO Logistic Regression remains a useful alternative because of its regularization, variable-selection capability, and interpretability, while achieving the same specificity as Random Forest.

The findings should, however, be interpreted cautiously because the analysis was based on a relatively small dataset and a single train-test partition. Future studies should evaluate the two approaches using larger and more diverse datasets, repeated or cross-validated resampling, and additional performance measures to determine whether the observed differences are robust and generalizable across different binary classification problems.

5.0 References

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1: Sensitivity and specificity. *BMJ, 308*(6943), 1552. https://doi.org/10.1136/bmj.308.6943.1552.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology, 110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004.

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics, 19*, 270. https://doi.org/10.1186/s12859-018-2264-5.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Gareth, J., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning: With Applications in R. Springer, Berlin, Heidelberg.

Gorman, R. P., & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks, 1*(1), 75–89. https://doi.org/10.1016/0893-6080(88)90023-8.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

https://doi.org/10.1007/978-0-387-84858-7.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153–157. https://doi.org/10.1007/BF02295996.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Declarations

Consent for Publication

Not applicable.

Availability of Data

The data used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no competing interests and no conflict of interest regarding the publication of this manuscript.

Ethical Consideration

Not applicable.

Funding

The authors declared no source of external funding.

Authors' Contributions

All components of the work were carried out by the author