

# Bayesian Inference in Breast Cancer Prediction

Yidong Wu

2024-06-16

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data Description</b>	<b>2</b>
<b>3</b>	<b>Exploratory Data Analysis</b>	<b>3</b>
3.1	Normalization . . . . .	3
3.2	Visualization of some Distributions . . . . .	3
3.3	Variable Selection with Correlation Heatmap . . . . .	5
<b>4</b>	<b>Model Descriptions</b>	<b>6</b>
4.1	Bayesian Logistic Regression . . . . .	6
4.2	Support Vector Machine . . . . .	7
4.3	Random Forest . . . . .	8
<b>5</b>	<b>Stan Code</b>	<b>8</b>
<b>6</b>	<b>Convergence Diagnostics</b>	<b>8</b>
6.1	Rhat . . . . .	8
6.2	Effective Sample Size . . . . .	9
6.3	Divergences . . . . .	9
<b>7</b>	<b>Posterior Predictive Checking</b>	<b>9</b>
<b>8</b>	<b>Predictive Performance</b>	<b>11</b>
8.1	Posterior Mean . . . . .	11
8.2	MAP . . . . .	11
8.3	Posterior Sampling . . . . .	11
8.4	Conclusion . . . . .	12
<b>9</b>	<b>Sensitivity Analysis</b>	<b>12</b>
<b>10</b>	<b>Model Comparison</b>	<b>14</b>
<b>11</b>	<b>Discussion and Further Improvements</b>	<b>15</b>
11.1	Cross Validation Approach . . . . .	15
11.2	Feature Selection Methods . . . . .	15
11.3	Prior and Parameter Choices . . . . .	15
11.4	Dataset Limitations and Advanced Models . . . . .	15
11.5	Conclusion . . . . .	15
<b>12</b>	<b>References</b>	<b>16</b>

# 1 Introduction

Breast cancer is a significant health concern, affecting approximately 2.3 million women annually worldwide and resulting in over 680,000 deaths each year. Early and accurate diagnosis is crucial for effective treatment and improving survival rates. Traditional diagnostic methods, such as clinical examinations and imaging, while essential, often suffer from limitations. These traditional approaches can be subjective, heavily dependent on the clinician's experience, and may lack strict probabilistic interpretation, leading to variability in diagnosis and potential for false positives and negatives.

In contrast, Bayesian methods offer a systematic approach that addresses these limitations. Bayesian inference incorporates prior knowledge and continuously updates predictions with new data, providing a probabilistic framework that quantifies uncertainty. This approach enhances prediction accuracy and offers reliable uncertainty estimates, making it particularly suitable for medical decision-making where certainty and precision are paramount.

In this project, we mainly apply Bayesian logistic regression to predict whether breast tumors are malignant or benign based on clinical and imaging features. By leveraging Bayesian methods, we aim to enhance the standard logistic regression model to incorporate prior knowledge and assess prediction uncertainty. Our goal is to improve diagnostic accuracy and reliability, supporting more informed clinical decisions.

## 2 Data Description

```
raw_data <- read.csv("breast_cancer.csv")
raw_data <- raw_data[, !(names(raw_data) %in% c("id", "X"))]
raw_data$diagnosis <- ifelse(raw_data$diagnosis == "M", 1, 0)
dim(raw_data)
```

```
## [1] 569 31
```

We will use the **Breast Cancer Wisconsin (Diagnostic)** dataset[1]. After dropping column **id** and **X**(empty column), there are 569 instances and 31 columns. The **diagnosis** column contains **M** and **B**, which means **malignant** and **benign**, and we use **1** and **0** to replace them respectively. The rest columns are features, computed from a digitized image of a fine needle aspirate (FNA) of a breast mass (see Figure 1). They describe characteristics of the cell nuclei present in the image. The **mean**, **standard error** and **worst** or largest (mean of the three largest values) of these features were computed for each image, resulting in 30 features.

Table 1: Description of Features in the Dataset

Feature	Description
radius	Mean of distances from center to points on the perimeter (Radius)
texture	Standard deviation of gray-scale values (Texture)
perimeter	Perimeter of the cell nuclei
area	Area of the cell nuclei
smoothness	Local variation in radius lengths (Smoothness)
compactness	Compactness (calculated as $\text{perimeter}^2 / \text{area} - 1.0$ )
concavity	Severity of concave portions of the contour (Concavity)
concave.points	Number of concave portions of the contour (Concave points)
symmetry	Symmetry of the cell nuclei
fractal_dimension	Fractal dimension (calculated as 'coastline approximation' - 1)

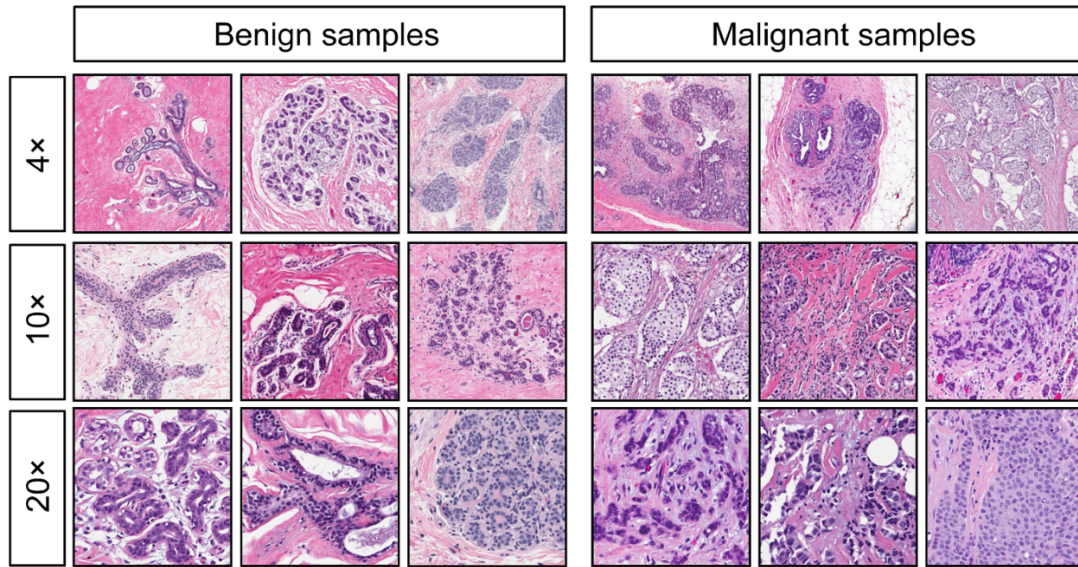


Figure 1: Image Samples

Next, we check the data set for missing values.

```
sum(is.na(raw_data))
```

```
## [1] 0
```

Great! There are no missing values in this dataset.

### 3 Exploratory Data Analysis

In this section, we conduct Exploratory Data Analysis to better understand the dataset. We will visualize some variables' distributions to identify patterns and trends. To aid in feature selection, we create a correlation heatmap to highlight highly correlated features, helping us to remove redundant features. This analysis provides essential insights for our Bayesian modeling approach.

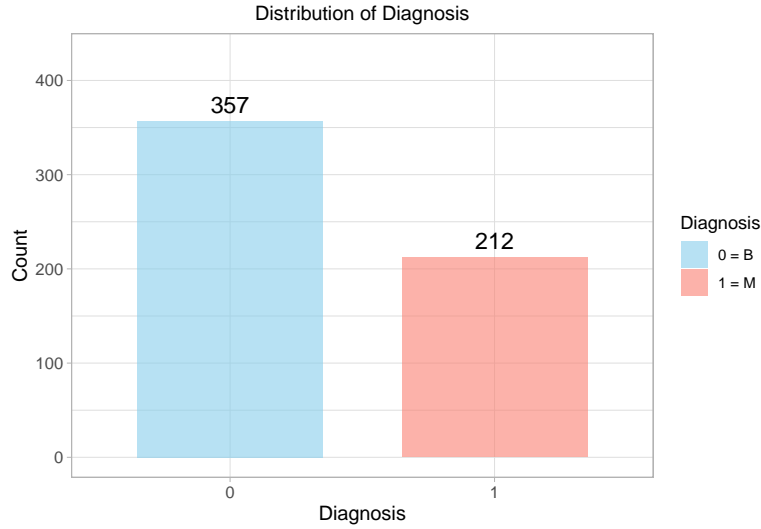
#### 3.1 Normalization

Normalization is essential to ensure that all features in the dataset contribute equally to the model. By rescaling features to a standard range, we eliminate differences in scales and units, which improves model performance and stability. This preprocessing step enhances the accuracy and reliability of our predictions.

```
diagnosis <- raw_data[, 1]
X <- raw_data[, -1]
X <- scale(X, center = TRUE, scale = TRUE)
normal_data <- data.frame(diagnosis,X)
```

#### 3.2 Visualization of some Distributions

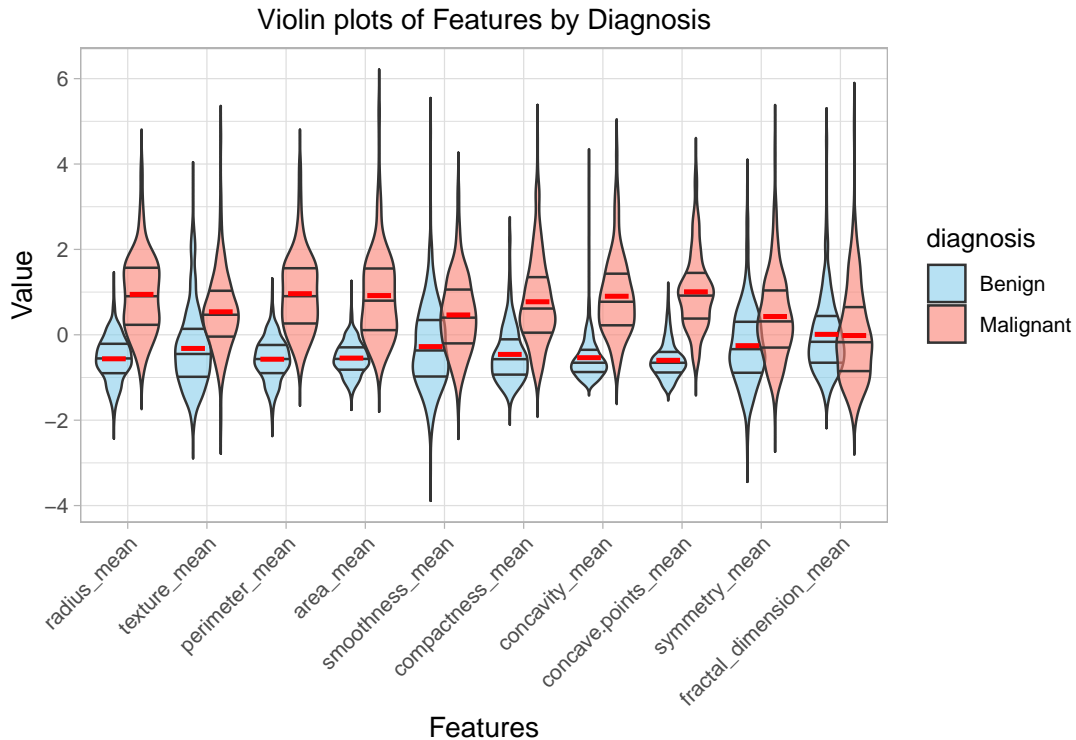
First, we examine the distribution of `diagnosis` in the dataset. The plot shows the count of benign (0) and malignant (1) diagnoses. We observe that there are 357 (62.74%) benign cases and 212 (37.26%) malignant cases, indicating an imbalance in the dataset. Specifically, we need to be careful about our evaluation metrics later on, such as precision, recall, and others, to ensure they appropriately reflect the model's performance given the class imbalance.



More specifically, considering the reality, the proportion of benign tumors is much higher than that of malignant tumors. So we need to use such metrics:

1. **Recall:** Recall indicates the proportion of actual malignant tumors that are correctly identified as malignant. This metric is critical because it reflects the model's ability to minimize false negatives, thereby reducing the likelihood of missing malignant tumors which could delay necessary treatment.
2. **Precision:** Precision represents the proportion of tumors identified as malignant that are indeed malignant. This is equally important as it ensures that the number of false positives is minimized, thereby avoiding unnecessary treatments and psychological stress for patients with benign tumors.

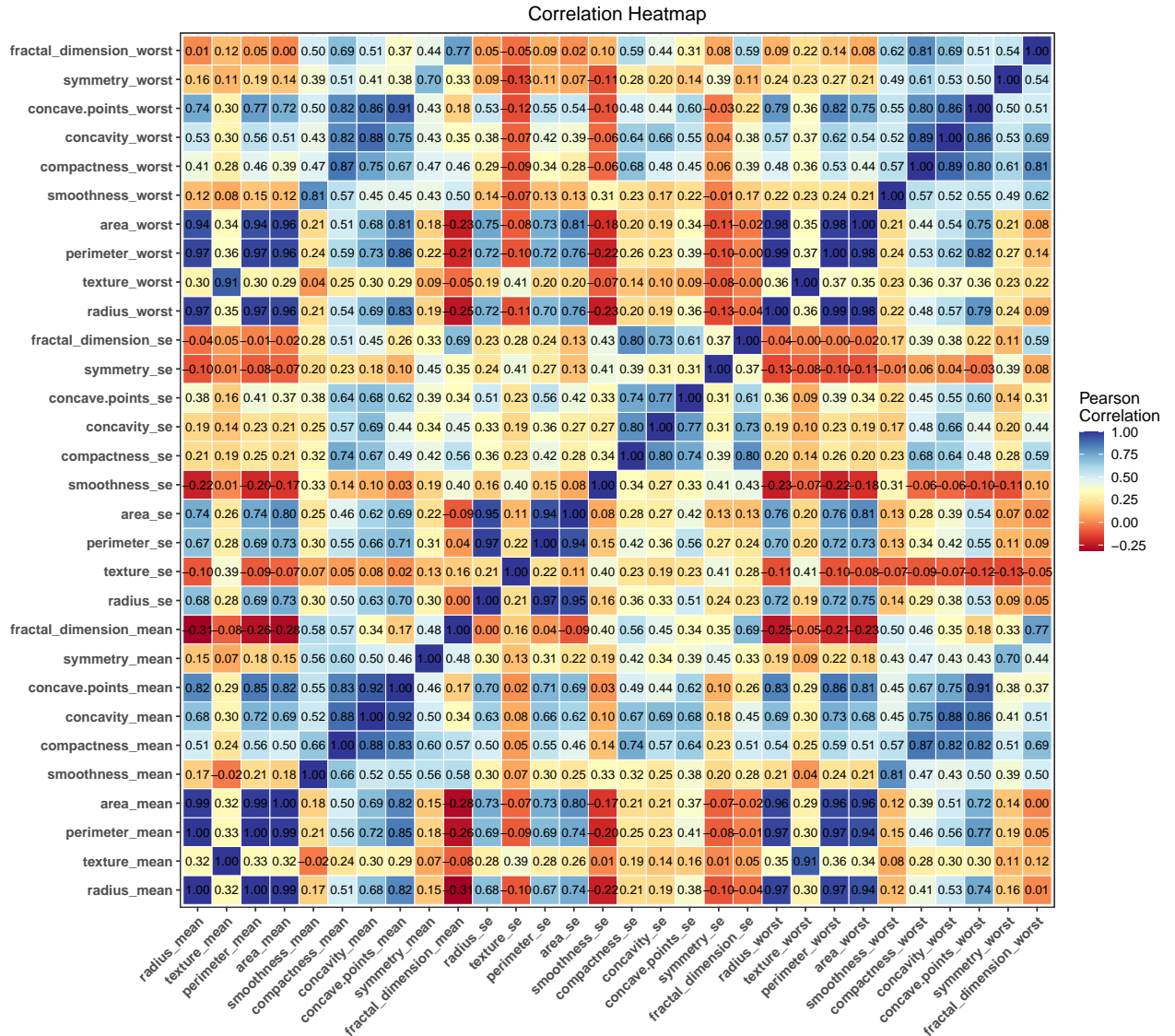
While **accuracy** is also an important metric, it is relatively less significant in the context of tumor detection. This is due to the typically higher prevalence of benign tumors compared to malignant ones. A high accuracy could potentially mask the model's inadequacy in detecting malignant tumors because the large number of correctly identified benign tumors can dominate the metric.



Next, we use violin plots to visualize the distribution of various features, stratified by diagnosis (benign vs. malignant). These plots provide a detailed view of the distribution shape, central tendency, and spread of each feature. The red line is the mean, and the three black lines are the quartiles. We observe distinct patterns and separations between benign and malignant cases for several features, indicating their potential importance in distinguishing between the two classes. Additionally, we note that the features are not skewed enough to warrant transformation. The violin plots for the remaining features exhibit similar patterns, so they are not shown here for simplicity. This visualization helps in understanding the variability and distribution of features, guiding the subsequent steps in feature selection and model development.

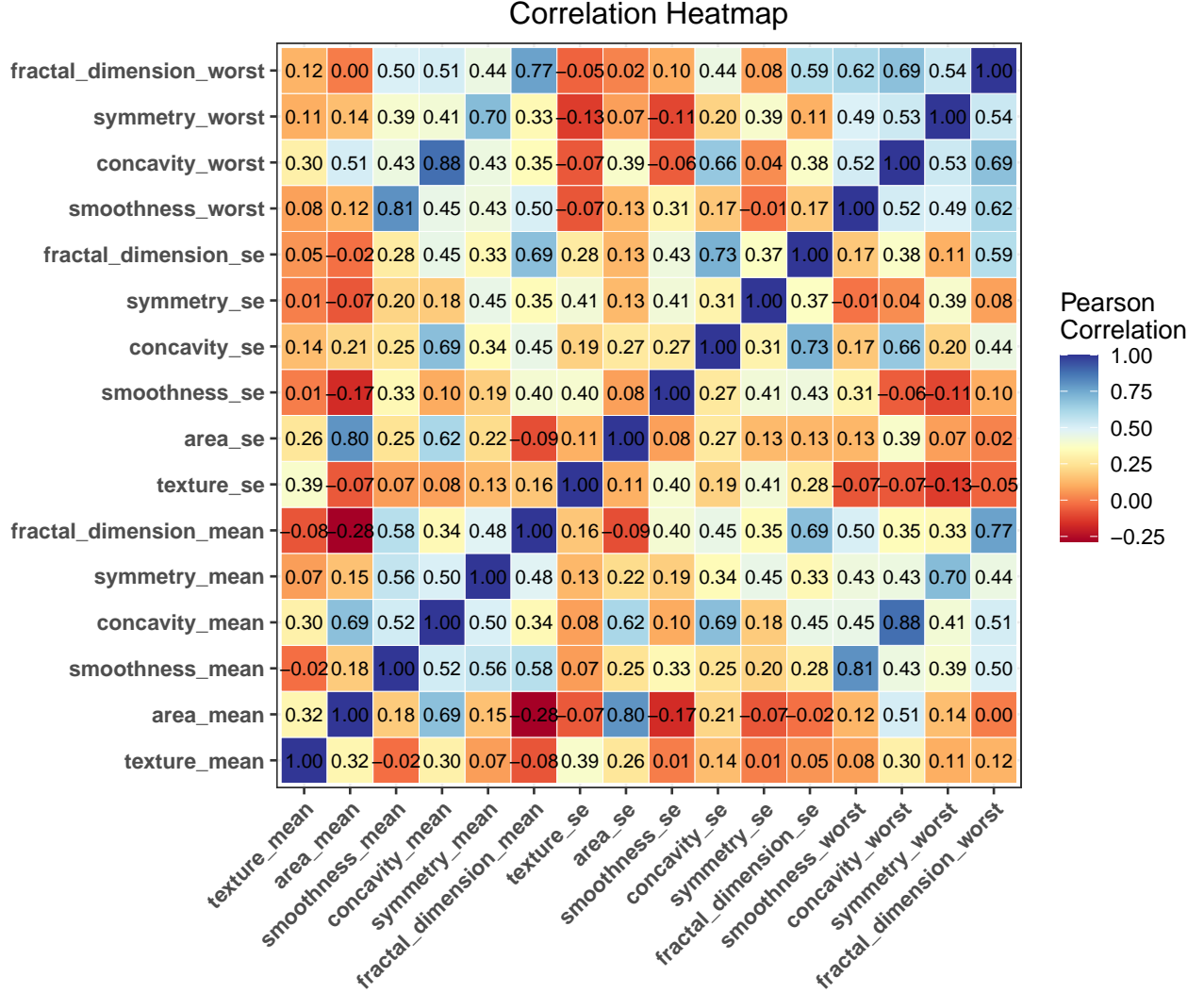
### 3.3 Variable Selection with Correlation Heatmap

We will use a correlation heatmap to identify and remove highly correlated features from the dataset, reducing multicollinearity and simplifying our model. After this preliminary selection, we fit the Bayesian logistic regression model using the remaining features. This approach ensures that we retain the most informative features, enhancing the model's accuracy and interpretability.



it may give more information. Note that there is no absolutely correct answer for variable selection. Finally, we choose the following features: `texture_mean`, `area_mean`, `smoothness_mean`, `concavity_mean`, `symmetry_mean`, `fractal_dimension_mean`, `texture_se`, `area_se`, `smoothness_se`, `concavity_se`, `symmetry_se`, `fractal_dimension_se`, `smoothness_worst`, `concavity_worst`, `symmetry_worst`, `fractal_dimension_worst`.

Then we check the correlation heatmap for them. We can see from the plot that the multicollinearity problem has been improved a lot.



## 4 Model Descriptions

### 4.1 Bayesian Logistic Regression

Bayesian Logistic Regression is a probabilistic approach to binary classification. In this model, we estimate the probability that a given instance belongs to a particular class. Let  $X = \{\vec{1}, X_1, X_2, \dots, X_p\}$  denote the set of predictor variables and  $y$  the binary response variable where  $y \in \{0, 1\}$ , with 0 representing benign and 1 representing malignant diagnoses.

The logistic regression model can be defined as:

$$\Pr(y = 1 \mid X, \beta) = \frac{1}{1 + \exp(-X\beta)}$$

where  $\beta = \{\beta_0, \beta_1, \dots, \beta_p\}$  are the coefficients to be estimated.

In the Bayesian framework, we place prior distributions on the coefficients  $\beta$ , which we will talk about later. The likelihood function for the logistic regression model is:

$$\mathcal{L}(y \mid X, \beta) = \prod_{i=1}^n \left( \frac{1}{1 + \exp(-X_i\beta)} \right)^{y_i} \left( 1 - \frac{1}{1 + \exp(-X_i\beta)} \right)^{1-y_i}$$

The posterior distribution of the coefficients  $\beta$  is then obtained by combining the prior and the likelihood:

$$p(\beta \mid X, y) \propto \mathcal{L}(y \mid X, \beta) \cdot p(\beta)$$

We can incorporate prior knowledge about the relationships between features and the diagnosis outcome, and obtain a full posterior distribution over the model parameters. This approach provides a measure of uncertainty in our estimates, which is particularly valuable for medical decision-making.

Next, we will choose different priors for this model. Since we don't have any prior information, we will explore three common types of priors: Gaussian, Laplace, and Cauchy. We assume  $\beta_j$  are i.i.d. for all these priors. Since we have standardized the data, we will refer to this in the selection of prior parameters. Additionally, we assume the prior for  $\alpha$  is  $\mathcal{N}(0, 10^2)$ . Here  $\alpha$  is actually the intercept  $\beta_0$ , since our  $X$  used in the following code does not have the intercept column.

#### 4.1.1 Gaussian Prior

The Gaussian prior is one of the most commonly used priors due to its convenient mathematical properties and interpretability. In our hierarchical model, we assume:

$$\begin{aligned} \beta_j \mid \theta, \sigma^2 &\sim \mathcal{N}(\theta, \sigma^2) \\ \theta \mid \sigma^2 &\sim \mathcal{N}(0, \frac{\sigma^2}{4}) \\ 1/\sigma^2 &\sim \text{Gamma}(4, 4) \end{aligned}$$

#### 4.1.2 Laplace Prior

The Laplace prior is commonly used in Lasso regression due to its property of inducing sparsity in the coefficient estimates. Here we assume:

$$\begin{aligned} \beta_j \mid b &\sim \text{Laplace}(0, b) \\ b &\sim \mathcal{N}(2, 1) \end{aligned}$$

#### 4.1.3 Cauchy Prior

The Cauchy prior is a heavy-tailed distribution that allows for occasional large coefficients, making it less restrictive than the Gaussian or Laplace priors. It is also proper[2]. Here we assume:

$$\begin{aligned} \beta_j \mid \gamma &\sim \text{Cauchy}(0, 2) \\ \gamma &\sim \mathcal{N}(2, 1) \end{aligned}$$

## 4.2 Support Vector Machine

Support Vector Machine (SVM) is a powerful supervised learning algorithm used for classification and regression tasks. SVM works by finding the hyperplane that best separates the data into different classes with the maximum margin. It is particularly effective in high-dimensional spaces and is robust to overfitting.



### 4.3 Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to improve classification or regression accuracy. Each tree in the forest is trained on a random subset of the data and features, and the final prediction is made by aggregating the predictions from all the trees. This approach reduces overfitting and enhances generalization, making Random Forest a versatile and widely used machine learning technique.

*The two machine learning methods mentioned above are only as a supplement and contrast, and this project is still mainly studying the model under the Bayesian framework.*

## 5 Stan Code

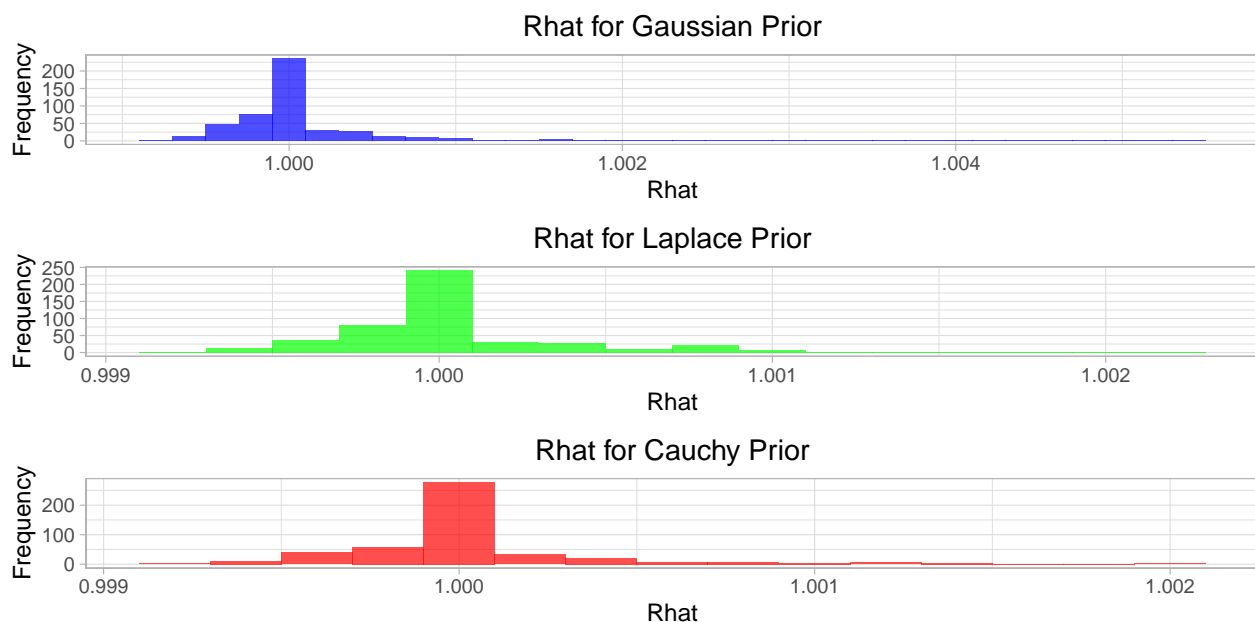
The specific Stan code for this section is too long to include in the main body of the report; it can be referenced in the accompanying R Markdown file. Briefly, we began by setting a random seed for reproducibility. We then split our dataset into training and test sets, using an 80/20 split. The training data was prepared for Stan by creating a list that includes the number of observations (N), the number of predictors (K), the diagnosis labels (y), and the predictor matrix (X).

We ran three different Bayesian logistic regression models with Gaussian, Laplace, and Cauchy priors, respectively. Each model was executed using the `stan` function, specifying the appropriate model code and data, with 5000 iterations across 4 chains to ensure robust sampling. This process allowed us to compare the performance of different priors in predicting whether a tumor is malignant or benign.

## 6 Convergence Diagnostics

### 6.1 Rhat

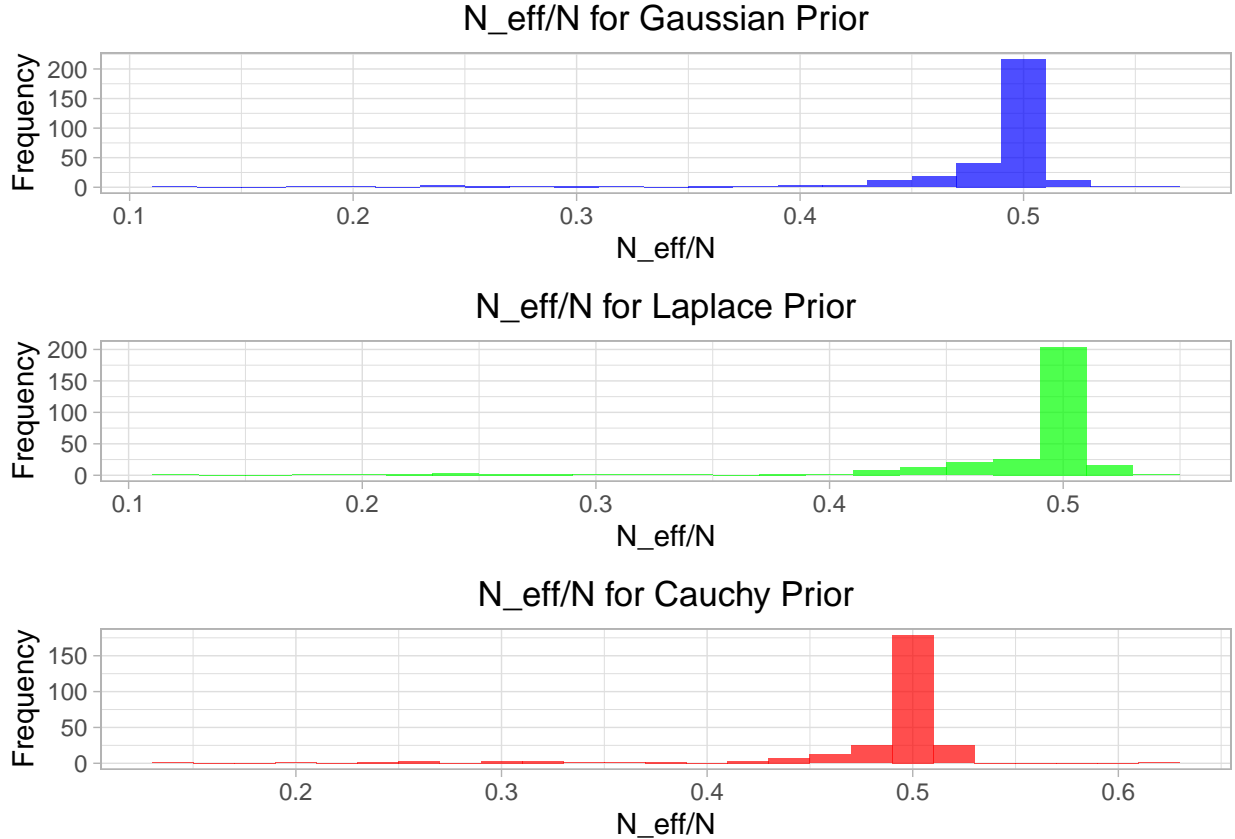
Rhat, also known as the Gelman-Rubin convergence diagnostic, is a measure used to assess the convergence of Markov Chain Monte Carlo (MCMC) simulations. From the plots, we can observe the Rhat values for all three models are very close to 1, signifying that the chains have converged adequately, and the MCMC simulations are reliable for inference.





## 6.2 Effective Sample Size

Here,  $N_{\text{eff}}$  represents the effective sample size, which estimates the number of independent samples in the MCMC chain, while  $N$  is the total number of samples drawn. For these three priors, the  $N_{\text{eff}}/N$  values are predominantly concentrated around 0.5, indicating that about half of the total samples can be considered effectively independent. This suggests a relatively good mixing of the chains and a lower degree of autocorrelation. However, the Cauchy prior presents a wider spread, which might indicate higher sensitivity to the data and potentially greater autocorrelation in some parameters.



## 6.3 Divergences

Divergences in MCMC sampling indicate potential problems where the sampler struggles to accurately explore the posterior distribution. In our analysis, we find zero divergences for all three priors. This indicates that the MCMC sampling process was stable and reliable across these priors, with no significant issues in exploring the posterior distributions.

```
## Number of divergences for Gaussian Prior: 0
```

```
## Number of divergences for Laplace Prior: 0
```

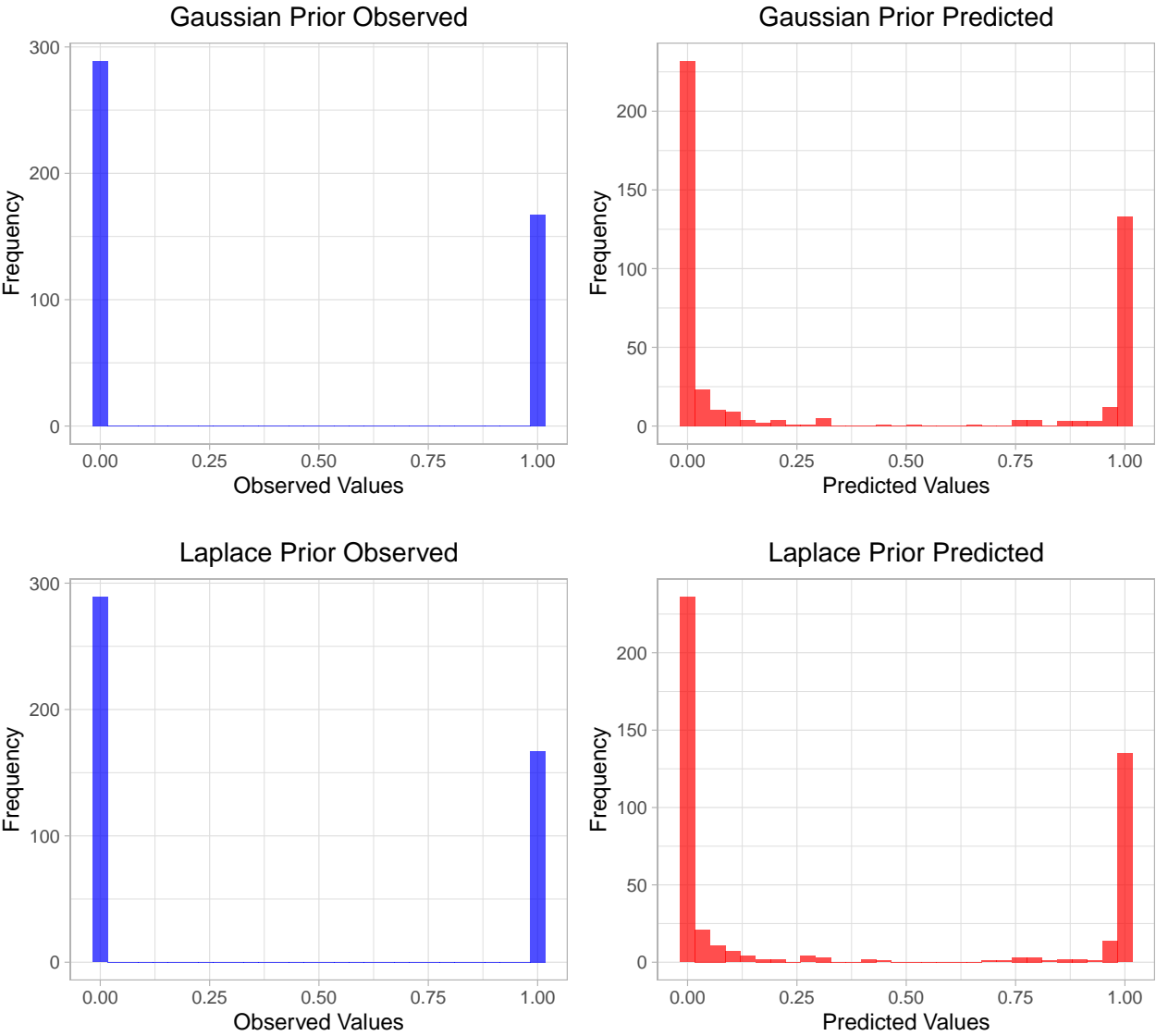
```
## Number of divergences for Cauchy Prior: 0
```

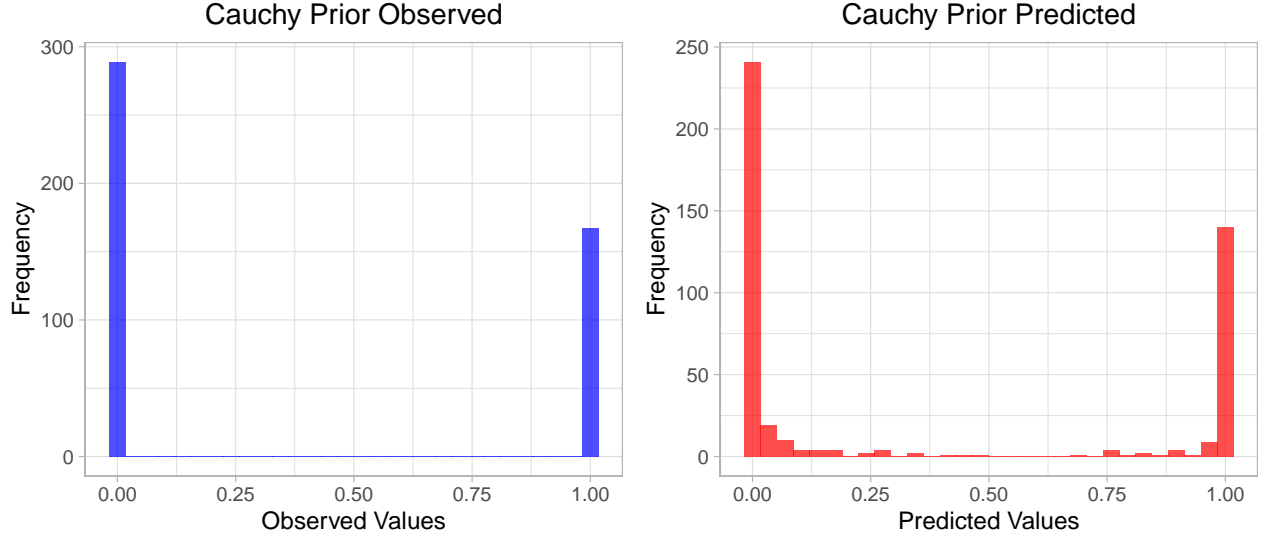
## 7 Posterior Predictive Checking

We compare the observed and predicted values (on training set) for our models with different priors: Gaussian, Laplace, and Cauchy. The histograms for the predicted values display not only the binary outcomes (0 and 1) but also a range of values in between. These intermediate values appear because we averaged the predictions across many samples. The sparsity of values around 0.5 suggests that the model is quite confident in its

predictions, seldom assigning probabilities that are close to indeterminate. In other words, the model tends to predict values that are very close to 0 or 1, indicating high confidence in the classification.

The model performance metrics—Precision, Recall, Accuracy, and F1 Score—are all very high, as shown in the table. It is important to note that these average metrics are calculated by first computing the metrics for each individual sample and then averaging them. This indicates that the models fit the training data exceptionally well. High precision means that when the model predicts a positive (malignant) class, it is usually correct. High recall indicates that the model captures most of the positive (malignant) class instances. The accuracy metric shows the proportion of total correct predictions, and the F1 Score, which is the harmonic mean of precision and recall, confirms the model’s robustness.





Model	Precision	Recall	Accuracy	F1 Score
Gaussian Prior	0.9637726	0.9635979	0.9576920	0.9632764
Laplace Prior	0.9693467	0.9691989	0.9608874	0.9693467
Cauchy Prior	0.9775595	0.9476506	0.9750747	0.9718035

## 8 Predictive Performance

### 8.1 Posterior Mean

In this section, we use the posterior means as the estimates for  $\alpha$  and  $\beta$  and run our models on the test set.

Model	Accuracy	Precision	Recall	F1 Score
Gaussian Prior	0.9646018	0.9571429	0.9852941	0.9710145
Laplace Prior	0.9646018	0.9571429	0.9852941	0.9710145
Cauchy Prior	0.9557522	0.9565217	0.9705882	0.9635036

### 8.2 MAP

In this section, we use the posterior modes as the estimates for  $\alpha$  and  $\beta$  and run our models on the test set.

Model	Accuracy	Precision	Recall	F1 Score
Gaussian Prior	0.9557522	0.9436620	0.9852941	0.9640288
Laplace Prior	0.9469027	0.9305556	0.9852941	0.9571429
Cauchy Prior	0.9646018	0.9571429	0.9852941	0.9710145

### 8.3 Posterior Sampling

In this section, we use all the posterior samples of  $\alpha$  and  $\beta$  to make predictions on the test set and then averaged the results.

Model	Accuracy	Precision	Recall	F1 Score
Gaussian Prior	0.9646018	0.9571429	0.9852941	0.9710145
Laplace Prior	0.9646018	0.9571429	0.9852941	0.9710145
Cauchy Prior	0.9557522	0.9565217	0.9705882	0.9635036

## 8.4 Conclusion

The predictive performance results reflect a few important insights about the choice of priors and estimation methods in Bayesian logistic regression. Firstly, the similarity in performance between the Gaussian and Laplace priors across all estimation methods (posterior means, MAP, and posterior sampling) suggests that these priors are robust choices for this dataset. Both priors yielded high accuracy, precision, recall, and F1 scores, indicating they are effective in capturing the underlying patterns in the data.

On the other hand, the Cauchy prior, while slightly underperforming in some cases, showed competitive results, especially when using the MAP estimates. This highlights the potential of the Cauchy prior in providing a different regularization effect that might be beneficial in certain scenarios, although it might not be as consistent as the Gaussian and Laplace priors.

The overall consistency of results across different estimation methods also suggests that the model’s performance is relatively stable regardless of whether we use posterior means, modes, or full posterior sampling. This stability is a positive indication that the Bayesian approach provides reliable and consistent predictions, which is crucial for practical applications.

## 9 Sensitivity Analysis

In the previous sections, we demonstrated that different priors (Gaussian, Laplace, and Cauchy) had a relatively minor impact on the final predictive performance. Now, we turn our focus to sensitivity analysis, specifically examining how variations in the parameters of the Gaussian prior influence the results.

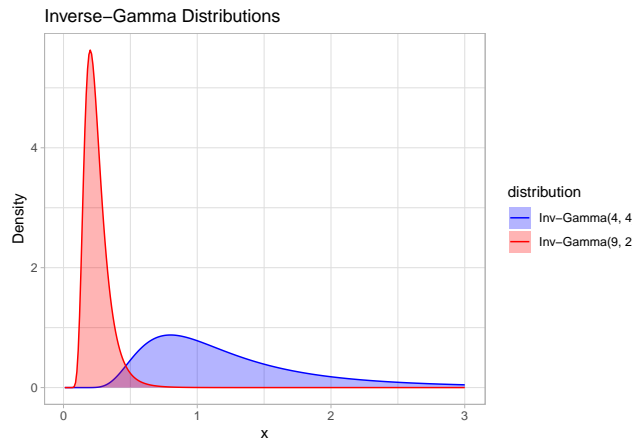
We will use these four Gaussian priors:

Prior 1:  $\alpha \sim \mathcal{N}(0, 10^2)$ ,  $\beta_j \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\theta \mid \sigma^2 \sim \mathcal{N}(0, \frac{\sigma^2}{4})$ ,  $1/\sigma^2 \sim \text{Gamma}(9, 2)$

Prior 2:  $\alpha \sim \mathcal{N}(0, 10^2)$ ,  $\beta_j \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\theta \mid \sigma^2 \sim \mathcal{N}(0, \sigma^2)$ ,  $1/\sigma^2 \sim \text{Gamma}(9, 2)$

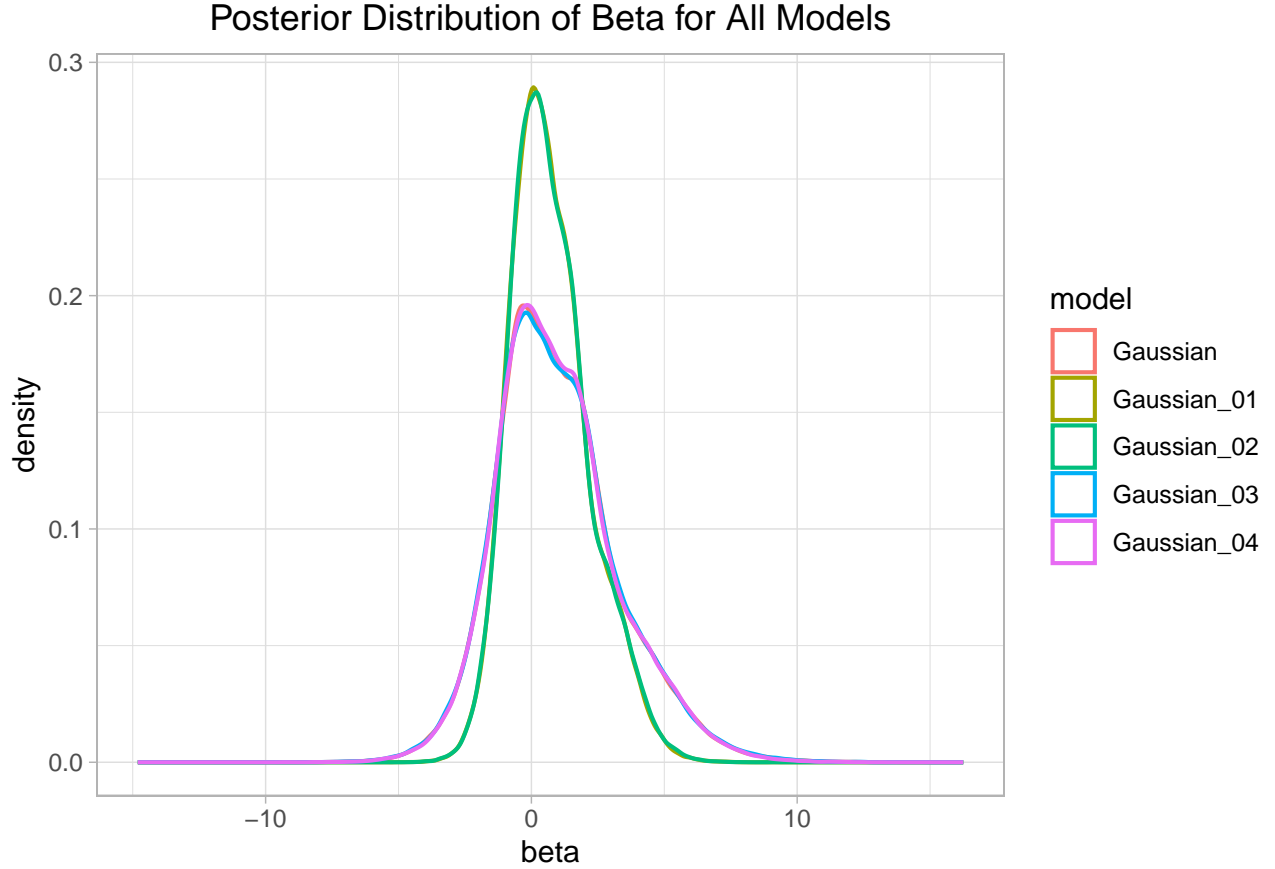
Prior 3:  $\alpha \sim \mathcal{N}(0, 100^2)$ ,  $\beta_j \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\theta \mid \sigma^2 \sim \mathcal{N}(0, \frac{\sigma^2}{4})$ ,  $1/\sigma^2 \sim \text{Gamma}(4, 4)$

Prior 4:  $\alpha \sim \mathcal{N}(0, 1^2)$ ,  $\beta_j \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ ,  $\theta \mid \sigma^2 \sim \mathcal{N}(0, \frac{\sigma^2}{4})$ ,  $1/\sigma^2 \sim \text{Gamma}(4, 4)$

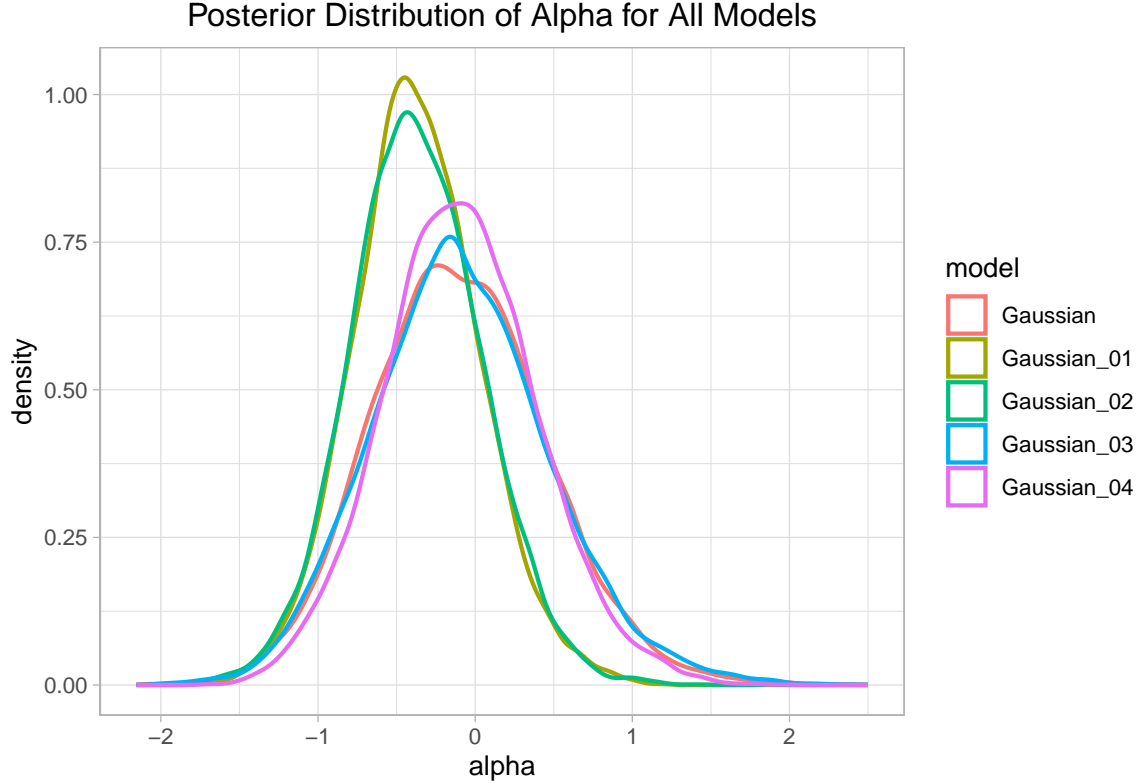


Following a similar approach, we can get the results:

Model	Accuracy	Precision	Recall	F1 Score
Gaussian Prior 0	0.9646018	0.9571429	0.9852941	0.9710145
Gaussian Prior 1	0.9557522	0.9436620	0.9852941	0.9640288
Gaussian Prior 2	0.9557522	0.9436620	0.9852941	0.9640288
Gaussian Prior 3	0.9646018	0.9571429	0.9852941	0.9710145
Gaussian Prior 4	0.9646018	0.9571429	0.9852941	0.9710145



From the first plot, we can observe that the distributions largely overlap, suggesting that the choice of prior does not drastically change the  $\beta$  estimates. The slight differences in the tail areas and the peaks could suggest that certain priors (Prior 1 and Prior 2) induce more regularization effects (smaller range for  $\sigma$ ) than others. In Prior 3 and Prior 4, we altered the prior distribution for  $\alpha$ . These changes did not lead to substantial differences in the posterior distribution of  $\beta$ , which is intuitive.



For  $\alpha$  in the second plot, changes in the prior had a slightly more noticeable impact. In Prior 1 and Prior 2, modifying the prior for  $\beta$  resulted in a significant shift in the posterior distribution of  $\alpha$ . This could be because  $\beta$  plays a dominant role in the logistic regression model, and changes in its prior can indirectly influence  $\alpha$  through their hierarchical relationship.

In Prior 3 and Prior 4, changing the prior distribution for  $\alpha$  did indeed affect its posterior distribution, but the changes were modest. This suggests that while the prior for  $\alpha$  has some influence, the data still play a dominant role in shaping the posterior distribution of  $\alpha$ .

The sensitivity analysis confirms that while  $\beta$  is relatively stable against changes in its prior,  $\alpha$  is more susceptible to prior modifications. But both parameters are generally robust overall. This indicates that the Bayesian logistic regression model is reliable, even when different priors are applied, making it a dependable approach for our analysis.

## 10 Model Comparison

Our ultimate goal is to achieve better model performance, such as higher recall and other metrics. In this section, we will compare the logistic regression models (with three different priors and without prior), SVM, and random forest using cross-validation to evaluate their performance. Given the high computational cost of leave-one-out (LOO) cross-validation, we will use k-fold cross-validation with  $k = 10$ .

Model	Accuracy	Precision	Recall	F1 Score
Gaussian	0.9683584	0.9689284	0.9801832	0.9743269
Laplace	0.9666040	0.9686110	0.9771035	0.9726442
Cauchy	0.9666040	0.9659043	0.9798062	0.9726817
No Prior LR	0.9613095	0.9649735	0.9714437	0.9680479
SVM	0.9630952	0.9659962	0.9744007	0.9699156
Random Forest	0.9630952	0.9629294	0.9779274	0.9701556

Here we add the traditional logistic regression model without a prior as a contrast. It can be seen that this model has the worst performance, indicating that the introduction of a prior is still effective. Actually, the Gaussian and Laplace prior can be viewed as ridge and lasso regularization (not exactly the same under our prior setup).

The Gaussian prior’s superior performance compared to the Laplace and Cauchy priors could be attributed to its more layered and complex structure, as well as the general ubiquity of the normality assumption. This suggests that the Gaussian prior is well-suited for capturing the underlying patterns in our data.

When it comes to SVM and Random Forest, despite their complexity, their performance did not surpass that of the logistic regression models. This may be due to the lack of fine-tuning of their respective parameters. For instance, we only employed the simplest linear kernel for SVM, and Random Forest might have performed better on a dataset with more raw features, as it is particularly effective at handling high-dimensional data and non-linear relationships.

Of course, these results are **not absolute**. There are other limitations in our analysis process that we will discuss in the next section. *However, as this is a Bayesian data analysis project, we are pleased to see that the Bayesian models performed the best, validating the value of our prior work and efforts.*

## 11 Discussion and Further Improvements

### 11.1 Cross Validation Approach

We used 10-fold cross-validation instead of Leave-One-Out (LOO) cross-validation, which might slightly affect the accuracy of our results. However, 10-fold cross-validation is a common and generally reliable method, balancing computational efficiency and performance. Further improvements can be accelerated using parallel computing.

### 11.2 Feature Selection Methods

In the Exploratory Data Analysis (EDA) section, we used a correlation heatmap to subjectively select some features. However, this approach lacks rigor. Future improvements could involve employing feature selection algorithms such as Recursive Feature Elimination (RFE) or decision tree-based methods. Alternatively, Principal Component Analysis (PCA) could be used for dimensionality reduction.

### 11.3 Prior and Parameter Choices

Our exploration of priors was limited, and the parameter choices might not be optimal. This is an inherent challenge, as finding the most appropriate priors often requires extensive experimentation. Moreover, we assume that  $\beta$  are i.i.d., which could not hold in reality. Additionally, we selected priors without historical data for reference, resulting in weakly informative priors. An improvement could be to use this dataset as historical data and apply our project methodology to a newer dataset.

### 11.4 Dataset Limitations and Advanced Models

Our dataset is relatively old, which may not fully reflect the latest medical advancements and diagnostic techniques. Additionally, it contains many redundant features that may not comprehensively represent the current situation. Future improvements could involve using advanced image processing technologies combined with Convolutional Neural Networks (CNN) to enhance diagnostic performance[3].

### 11.5 Conclusion

Overall, while our Bayesian data analysis project has yielded valuable insights, there are several areas for further refinement and improvement. We believe that the Bayesian framework can be applied to many data analysis practices to bring better results.



## 12 References

1. UCI Machine Learning Repository. (2023). Breast Cancer Wisconsin (Diagnostic) Data Set. Available at: <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data>
2. Ghosh, Joyee. "Cauchy and other shrinkage priors for logistic regression in the presence of separation." *Wiley Interdisciplinary Reviews: Computational Statistics* 11.6 (2019): e1478.
3. Sharma, Anuraganand, and Dinesh Kumar. "Classification with 2-D Convolutional Neural Networks for breast cancer diagnosis." *Scientific Reports* 12.1 (2022): 21857.
4. Hoff, Peter D. *A first course in Bayesian statistical methods*. Vol. 580. New York: Springer, 2009.
5. Gelman, Andrew, et al. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.