# Enhancing Breast Cancer Diagnosis:
# Integrative Approaches with Bayesian Methods and Machine Learning

## 1. Introduction

This proposal conducts a binary classification problem that is used to diagnose breast cancer. The dataset is Breast Cancer Dataset. The input is 29 different features of the patient, and the output is malignant or benign. Also, we discuss the sensitivity analysis and compare the performance of Bayesian model with machine learning models. (Code and result figures: https://github.com/Yuan-Kivi/engg-107-final-project)

## 2. Literature Review

Banerjee (2020) explores the application of Bayesian networks to predict breast cancer risk [1]. Aloraini (2019) highlights the use of machine learning techniques for predicting breast cancer [2]. Although there are also some state-of-the-art models like XBNet and MONT3, there is still no one comparing the Bayesian model and machine learning models together to see their difference of performances [3] [4]. Thus, this is the contribution of this proposal.

## 3. Hypothesis and Expected Result

Bayesian method is suitable for binary classification problem and can acquire excellent result compared with other machine learning algorithm. To make the hypothesis testable, I will compare accuracy, recall and precision and see whether the difference of Bayesian model and the best machien learning model is $\pm$ 2.5%.

## 4. Bayesian Model

There are many approaches in Bayesian modeling like pre-calibration, Bayesian Monte Carlo (BMC), Markov Chain Monte Carlo (MCMC). In this proposal, I use MCMC with Metroplis-Hasting algorithm as the Bayesian model. The reason is it can handle high dimensional problem because we can optimize to approach the optimal parameter vector instead of computing the whole posterior distribution.

### 4.1. **Model.**

$$f(x, \theta) = \text{sigmoid}(\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N)$$

$$y_{\text{obs}} = f(x, \theta) + \varepsilon$$

Here, N is 29 because each patient has 29 features information.

4.2. **Prior.** I use Cauthy distribution with center 0 and scale 2.5 as the prior distribution. Andrew Gelman and his colleagues (2008) find $Cauthy(0, 2.5)$ is very suitable as the prior knowledge in logistic regression [5].

$$\text{Cauchy}(0, 2.5) = P(\theta) = \frac{1}{2.5\pi\left[1 + \left(\frac{x}{2.5}\right)^2\right]}$$

4.3. **Likelihood.** In binary classification problem, Bernoulli distribution is suitable for the likelihood function.

$$L(\theta|X,y) = \prod_{i=1}^{N} p(y_i|x_i;\theta) = \prod_{i=1}^{N} \left(\frac{1}{1+e^{-x_i^T\theta}}\right)^{y_i} \left(1 - \frac{1}{1+e^{-x_i^T\theta}}\right)^{1-y_i}$$

Here, $N$ is the number of patients. The total number of the dataset is 570. I divide into 80% as training dataset so $N$ should be 456.

4.4. **Posterior.**

$$p(\theta|X,y) \propto L(\theta|X,y) \times p(\theta)$$

The posterior is proportional with prior multiple likelihood.

4.5. **Convergence.** I set the iteration time is 100000 and I use R-hat diagnosis and Effective sample size (ESS) to test the convergence. According to the result, R-hat is 1.053 which is closed to 1 and not greater than 1.1. It means the model convergence [6]. ESS is 954500 which is related to the total sample size, so it shows the convergence as well.

4.6. **Sensitivity Analysis.** I use sobol method to do sensitivity analysis. I set the $N$ is 1024 and parameter bound is $[-20, 20]$. The result shows the first five contribution features are radius stand error, perimeter stand error, concave points mean, fractual dimension worst and perimeter worst. The further research can concentrate on these five features.

## 5. Machine Learning Approaches

I use Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Decision Tree (DT) to do the binary classification [7] [8] [9].

## 6. Evaluation

According to the confusion matrix, the accuracy are 95.61% (MCMC), 99.12% (MLP), 97.36% (SVM), 88.59% (DT). The recall are 90.69% (MCMC), 97.67% (MLP), 93.02% (SVM), 83.72% (DT). The precision are 97.5% (MCMC), 100% (MLP), 100% (SVM), 85.71% (DT).

## 7. Conclusion

MCMC with M-H algorithm acquire a 95.61% accuracy, 90.69% recall and 90.75% precision in testing dataset. It is better than decision tree, the same as support vector machine and less than 2.5% than MLP. It gives the evidence of our hypothesis.

## 8. Intellectual Merit and Broader Impacts

We use MCMC methods handle this complex model by efficiently sampling from the posterior distributions of model parameters and allow for the explicit incorporation of this prior knowledge in the analysis, enhancing the interpretability and relevance of the results. Meanwhile, we can do model comparison. Last, the sensitivity analysis shows there are five features contibute the most of prediction which can lead the following research concentrating in those features. Breast cancer is one of the most fatal diseases nowadays. The most effective way to cure cancer is by diagnosing it at an early stage. However, the normal way to detect breast cancer is laborious and will cause the patient to suffer a lot of pain. Thus, using the Bayesian model to diagnose breast cancer will make the process easier.

## References


[1] Banerjee, S., & Mitra, S. (2020). Application of Bayesian network in breast cancer risk prediction. International Journal of Healthcare Information Systems and Informatics (IJHISI), 15(1), 63-76.

[2] Aloraini, A. (2019). Breast cancer prediction using machine learning. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 1-4. IEEE.

[3] Tushar Sarkar, XBNet: An Extremely Boosted Neural Network. arXiv:2106.05239v3 [cs.LG] 21 Apr 2022.

[4] Varun Ojha, Giuseppe Nicosia. Multi-objective Optimisation of Multi-output Neural Trees. arXiv:2010.04524v2 [cs.NE] 18 Feb 2022.

[5] Andrew Gelman,Aleks Jakulin, Maria Grazia Pittau, Yu-Sung Su. (2008) A weakly informative default prior distribution for logistic and other regression models

[6] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved Rb for assessing convergence of MCMC

[7] Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function Mathematics of Control, Signals, and Systems, 2(4), 303–314.

[8] Cortes, Corinna; Vapnik, Vladimir (1995). "Support-vector networks" (PDF). Machine Learning. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362. doi:10.1007/BF00994018. S2CID 206787478

[9] von Winterfeldt, Detlof; Edwards, Ward (1986). "Decision trees". Decision Analysis and Behavioral Research. Cambridge University Press. pp. 63–89. ISBN 0-521-27304-8.