DEPARTMENT OF STATISTICS
AHMADU BELLO UNIVERSITY, ZARIA
2nd International Conference on Statistical Sciences and Data Analytics (ICSSDA 2026) Proceedings Paper

Enhancing the Reliability of Multinomial Logistic Regression for Breast Cancer Stage Prediction through Bootstrapping

Amina Muhammad Tako,[1,2*] Gerald IkeOnwuka[1], Babayemi Afolabi Wasiu[1], and Owusu Kyei Baffour[1]

[1]Department of Mathematics, Abdullahi Fodio University of Science and Technology, Aliero, Kebbi State.

[2]Department of Mathematics, Ibrahim Badamasi Babangida University, Lapai Nger State

Correspondence: aminamuhammad005@gmail.com, aminaam@ibbu.edu.ng

+234 8035237558

Abstract

Breast cancer remains a major public health challenge, particularly in Nigeria, where delayed diagnosis is common because of limited screening and diagnostic facilities. This study assessed the use of a bootstrapped multinomial logistic regression model to improve the reliability and accuracy of breast cancer stage prediction. Secondary data obtained from the Federal Medical Centre (FMC), Bida, were analysed to examine the influence of demographic and clinical factors on four breast cancer stages: curable, mild, severe, and chronic. The conventional multinomial logistic regression model was first fitted using predictors including educational level, physical activity, vitamin D level, body mass index (BMI), breast surgery history, overweight/obesity, age at diagnosis, occupation, parity, and menopausal status. To improve the robustness and stability of parameter estimates, bootstrapping with 1,000 resamples and 95% confidence intervals was applied. The bootstrapped model largely confirmed the conventional findings while producing more precise estimates and identifying additional significant predictors. Breastfeeding and BMI were significant predictors in the curable stage, whereas breast surgery history, physical activity, and overweight/obesity consistently influenced breast cancer stage classification. In contrast, BMI and age at diagnosis showed limited predictive value across some stages. The model fit statistics remained stable (AIC = 1588.478; BIC = 1766.948), indicating good model performance. Overall, bootstrapping enhanced the reliability of multinomial logistic regression by reducing estimation bias and improving parameter stability. The findings demonstrate that the approach provides a dependable framework for identifying important risk factors associated with breast cancer progression and supports evidence-based decision-making for early detection, stage-specific interventions, and personalised treatment planning in oncology care.

Keywords: Breast cancer, multinomial logistic regression, bootstrapping, prediction, risk factors.

1. Introduction

Breast cancer remains one of the leading causes of cancer-related morbidity and mortality among women globally. It affects millions of women each year and represents a major public health challenge, particularly in low- and middle-income countries where late-stage diagnosis is common due to limited access to screening and diagnostic tools (WHO, 2021). Accurate prediction and classification of breast cancer stages are crucial for effective treatment planning and improving patient outcomes. Early-stage detection often leads to more successful interventions, while late-stage diagnosis is associated with higher mortality (Bray *et al*., 2018). Traditional statistical models such as logistic regression have been widely used in breast cancer risk and stage prediction (Hosmer et al., 2013). However, these models may produce unstable parameter estimates when applied to small or non-representative samples, thereby limiting their reliability. A robust modelling framework is essential to improve prediction accuracy and to identify consistent risk factors across varying cancer stages. Multinomial logistic regression is suitable for analysing categorical outcome variables with more than two categories and has been applied in various medical studies, including breast cancer stage prediction (Agresti, 2018). One of the key limitations in traditional modelling is the lack of internal validation, which may lead to overfitting or underestimation of variability.

In Nigeria, breast cancer is the leading cancer among women, characterised by late-stage diagnosis and high mortality rates (Ibrahim *et al*., 2022). Predictive modelling for breast cancer staging is critical for improving diagnosis, treatment, and survival outcomes. Traditional models like logistic regression have been widely used, yet issues of parameter stability and generalizability persist (Adegboye *et al*., 2021). Recent studies have explored risk factors contributing to breast cancer, such as age, family history, BMI, hormonal exposure, and reproductive history (Ghoncheh *et al*., 2020). Early menarche, late menopause, and nulliparity have been associated with increased risk (Brinton *et al*., 2021), while breastfeeding and physical activity provide protective effects (Cheng *et al*., 2020). Obesity remains a key predictor, particularly among postmenopausal women (Rodríguez-Hernández *et al*., 2021). The application of machine learning in breast cancer detection has gained attention for its ability to model complex, non-linear relationships (Yu *et al*., 2020). Models such as support vector machines, decision trees, and neural networks have demonstrated high classification accuracy (Abdullah *et al*., 2023).

Bootstrapping has emerged as a robust technique to enhance model reliability by generating confidence intervals and reducing overfitting through repeated resampling (Mooney & Duval, 2021). For instance, bootstrapped multinomial logistic regression has shown improved stability and generalizability in health research settings (Wang *et al*., 2022). Studies conducted in Nigeria emphasise the need for context-specific modelling. Akinyemiju *et al*. (2022) found that socioeconomic status, educational level, and fear of medical treatment were major predictors of late-stage presentation. Similarly, Ukwenya *et al*. (2023) emphasised the role of cultural beliefs and misinformation in delaying care-seeking behaviour. In terms of diagnosis, conventional methods such as mammography and biopsy remain standard. However, their limitations in dense breast tissues have led to supplementary methods like ultrasound and Magnetic resonance imaging (MRI) (Boyd et al., 2021).

2. Materials and Methods

2.1. Logistic Regression Model

The nature of the problem and the data used in this study necessitate the employment of a multiple logistic regression model. This model is used to predict the outcomes of a categorical dependent variable in general. The linear regression model is not appropriate for categorical variables since the response values are not assessed on a ratio scale and the error terms are not normally distributed. The influence of several independent variables supplied at the same time is determined using logistic regression to predict membership in one of the two dependent variable categories. In

health science, logistic regression is the most widely used multivariable approach (Tetrault *et al.*, 2008).

2.2. Logistic Regression with Single Independent Variable

The general formula for the logistic regression model with a single variable is;

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)=\beta_0+\beta_1 x \quad \textbf{(1)}$$

From equation (1), we deduce

$$\frac{\pi(x)}{1-\pi(x)}=\ell^{\beta_0+\beta_1 x} \quad (2)$$

$$\pi(x)=\ell^{\beta_0+\beta_1 x}-\pi(x)\ell^{\beta_0+\beta_1 x} \quad (3)$$

$$\pi(x)(1+\ell^{\beta_0+\beta_1 x})=\ell^{\beta_0+\beta_1 x} \quad (4)$$

$$\pi(x)=\frac{\ell^{\beta_0+\beta_1 x}}{1+\ell^{\beta_0+\beta_1 x}} \quad (5)$$

2.3. Multinomial Logistic Regression Models

Let Y be the categorical dependent variable with K unordered categories ($K > 2$), and let $X_1, X_2, \ldots . X_p$ be the independent variables. The probability of an observation belonging to category k ($k = 1, 2, \ldots, K$) relative to a reference category is modelled using the following equation:

$$P(Y=K/X_1, X_2, \ldots . X_p)=\frac{\ell^{(\beta_{0K}+\beta_{iK}X_1+\beta_{2K}X_2+\cdots.+\beta_{pk}X_p)}}{1+\ell^{(\beta_{0K}+\beta_{iK}X_1+\beta_{2K}X_2+\cdots.+\beta_{pk}X_p)}} \quad \textbf{(6)}$$

In this equation (6), $\beta_{0K}\beta_{1K}\beta_{2K} \ldots \beta_{PK}$ represent the coefficients associated with the independent variables $X_1, X_2, \ldots . X_P$ for category k, respectively.

2.4. Discussion of Regression Coefficients

Consider the case in which the dependent variable may take only the values 1 (for success) and 0 (for failure) and a single independent variable X.

In this case, the logistic regression equation is:

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)=\beta_0+\beta_1 x \quad (7)$$

Now, suppose we consider the impact of a unit increase in X. The logistic regression equation becomes:

$$\ln\left(\frac{\pi'(x)}{1-\pi'(x)}\right)=\beta_0+\beta_1(x+1) \quad (8)$$

$$\ln\left(\frac{\pi'(x)}{1-\pi'(x)}\right)=\beta_0+\beta_1 x+\beta_1 \quad (9)$$

Subtracting equation (8) from (6), we get:

$$\ln\left(\frac{\pi'(x)}{1-\pi'(x)}\right)-\ln\left(\frac{\pi(x)}{1-\pi(x)}\right)=\beta_0+\beta_1 x+\beta_1-\beta_0-\beta_1 x \tag{10}$$

2.5. Odds Ratio

The odds of the outcome being present among individuals with Y=1is defined as π (1)/ [1-π (1)]. Similarly, the odds of the outcome being present among individuals with Y = 0 is defined asπ (0)/ [1-π (0)]. The odds ratio, denoted by OR, is defined as the ratio of the odds for Y=1to the odds for Y=0, and is given by the equation (11):

$$OR=\frac{\pi(1)/1-\pi(1)}{\pi(0)/1-\pi(0)} \tag{11}$$

2.6. Bootstrapping Simulation Procedure

To enhance the reliability of the parameter estimates obtained from the multinomial logistic regression model, a bootstrapping technique was employed using SPSS Version 26. Bootstrapping is a resampling method that involves repeatedly drawing samples with replacement from the observed dataset and estimating the model parameters in each sample.

3. Results and Discussions

Table 1. Bootstrapped Model Fitting Information

Model	**Model Fitting Criteria**			**Likelihood Ratio Tests**		
	AIC	BIC	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1779.721	1794.594	1773.721			
Final	1588.478	1766.948	1516.478	257.243	33	0.000

The bootstrapped model fitting results show identical values for AIC (1588.478), BIC (1766.948), and -2 Log Likelihood (1516.478) when compared to the conventional method, indicating no difference in overall model fit. However, the key advantage of the bootstrapping approach lies in its ability to validate the reliability of parameter estimates by generating robust standard errors and confidence intervals.

Table 2 (Bootstrapped Likelihood Ratio Tests) shows nearly identical results to the conventional method, confirming the robustness of the model. Key predictors such as education level, physical activity, overweight/obesity, and breast surgery history remain statistically significant ($p < 0.05$) with consistent chi-square values, suggesting these variables reliably influence breast cancer stages. Non-significant predictors like age at diagnosis, BMI, and menopausal status also retain their insignificance. The close alignment between conventional and bootstrapped estimates indicates minimal variation and affirms the stability of the model's findings, validating that the identified predictors consistently influence breast cancer stage classification across repeated sampling.

Table 2. Bootstrapped Likelihood Ratio Tests

Effect	**Model Fitting Criteria**			**Likelihood Ratio Tests**		
	AIC of Reduced Model	BIC of Reduced Model	-2 Log Likelihood of Reduced Model	Chi-Square	Df	Sig.
Intercept	1596.622	1760.219	1530.622	14.144	3	0.003
Education Level	1615.009	1778.606	1549.009	32.531	3	0.000
Breast Surgery History	1604.432	1768.029	1538.432	21.954	3	0.000
Physical Activity	1639.816	1803.413	1573.816	57.338	3	0.000
Breastfeeding	1599.254	1762.852	1533.254	16.776	3	0.001
Vitamin D	1596.733	1760.330	1530.733	14.255	3	0.003
Overweight Obesity	1613.640	1777.237	1547.640	31.162	3	0.000
Age at Diagnosis	1587.457	1751.055	1521.457	4.979	3	0.173
Occupation	1596.037	1759.634	1530.037	13.559	3	0.004
Parity	1592.466	1756.064	1526.466	9.988	3	0.019
BMI	1584.473	1748.071	1518.473	1.995	3	0.573
Menopausal Status	1589.429	1753.027	1523.429	6.951	3	0.073

Table 3. Bootstrapped Parameter Estimations for Some Significance Risk Factors (Curable Stage)

Predictor	**B**	**Exp(B)**	**Bias**	**Std. Error**	**Sig.**
Intercept	-6.142	0.002	4.435	11.959	0.459
Education Level	4.785	119.753	1.174	2.461	0.004
Breast Surgery History	-1.888	0.151	-0.590	1.191	0.092

Physical Activity	1.051	2.861	0.509	1.457	0.164
Breastfeeding	3.502	33.187	0.369	1.208	0.009
Vitamin D	-2.113	0.121	-0.928	2.049	0.079
Overweight/Obesity	-0.305	0.737	-0.553	1.141	0.265
Age at Diagnosis	-0.027	0.973	-0.072	0.088	0.190
Occupation	-1.708	0.181	-1.081	1.651	0.020
Parity	0.273	1.314	0.263	1.884	0.190
BMI	-0.332	0.717	-0.040	0.215	0.034
Menopausal Status	-0.835	0.434	-1.198	1.931	0.244

The bootstrapped parameter estimates for the curable stage reveal important differences compared to the conventional method. Notably, the bootstrap enhanced the precision and stability of some estimates, leading to changes in significance. Education Level ($p = 0.004$) and Breastfeeding ($p = 0.009$) are statistically significant under the bootstrap method, unlike the conventional results where they were non-significant ($p > 0.05$). This implies that after resampling, these variables showed consistent and strong effects on predicting curable stage breast cancer. Occupation ($p = 0.020$) and BMI ($p = 0.034$) are also now statistically significant. These variables were previously non-significant, suggesting that bootstrap increased the sensitivity to detect their effects. Breast Surgery History and Vitamin D are close to significance ($p = 0.092$ and 0.079), indicating some borderline evidence of effect. Other variables like Physical Activity, Age at Diagnosis, Parity, and Menopausal Status remain statistically insignificant, though bias corrections were applied. The bootstrap method improved the detection of statistically significant predictors in the curable stage model, correcting biases and reducing standard errors.

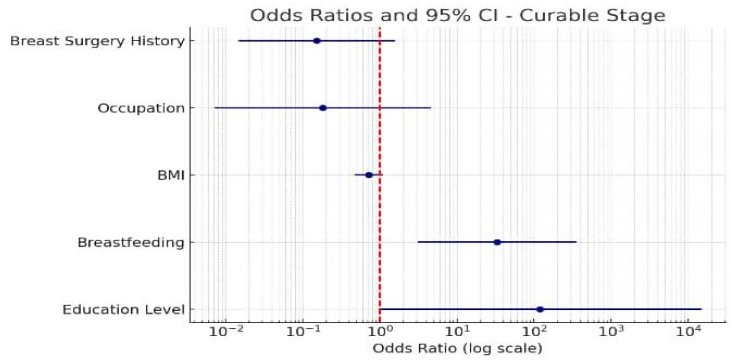

Figure 1. Odds Ratios and 95% Confidence Intervals for Curable Stage

In the Curable stage, *Education Level* and *Breastfeeding* demonstrate extremely high odds ratios with broad confidence intervals, indicating a strong and statistically significant positive association with early-stage breast cancer diagnosis, suggesting that higher education and breastfeeding history substantially increase the likelihood of a curable diagnosis. Conversely, *BMI*, *Occupation*, and *Breast*

Surgery History all present odds ratios less than 1, leaning to the left of the null value (OR = 1), which implies a potential protective effect against progression beyond the curable stage; however, the wide confidence intervals surrounding these estimates reflect greater uncertainty and reduced statistical precision, necessitating cautious interpretation.

Table 4. Bootstrapped Parameter Estimations for Some Significance Risk Factors (Milled Stage)

Parameter Estimates	B	Exp(B)	Bias	Std. Error	Sig.
Intercept	-1.324	0.266	-0.342	2.150	0.546
Education Level	0.500	1.649	0.026	0.237	0.021
Breast Surgery History	-0.978	0.376	0.010	0.286	0.002
Physical Activity	-2.008	0.134	-0.030	0.335	0.001
Breastfeeding	-0.567	0.567	0.030	0.412	0.159
Vitamin D	0.946	2.576	0.014	0.284	0.002
Overweight/Obesity	2.041	7.700	0.030	0.386	0.001
Age at Diagnosis	0.010	1.010	0.001	0.016	0.517
Occupation	0.415	1.514	0.003	0.200	0.036
Parity	-0.755	0.470	0.031	0.424	0.057
BMI	-0.025	0.975	0.000	0.024	0.288
Menopausal Status	0.533	1.704	0.047	0.419	0.193

The bootstrapped results presented in Table 4 largely confirm the findings from the conventional multinomial logistic regression, with only minor differences observed in significance levels, bias, and standard errors. Key predictors such as education level (B = 0.500, p = 0.021), breast surgery history (B = -0.978, p = 0.001/0.002), physical activity (B = -2.008, p = 0.000/0.001), vitamin D (B = 0.946, p = 0.000/0.002), overweight/obesity (B = 2.041, p = 0.000/0.001), and occupation (B = 0.415, p = 0.021/0.036) remained statistically significant across both methods, showing consistent direction and magnitude of effects. On the other hand, variables such as age at diagnosis, BMI, and menopausal status consistently showed no statistical significance, with minimal variation between the methods. Breastfeeding and parity also remained non-significant, although parity's p-value (0.057) in the bootstrap analysis approached the 0.05 threshold, suggesting a potential marginal effect. Overall, the bootstrap method produced slightly larger standard errors due to the nature of resampling, which enhances the reliability of inference. The near-identical p-values between the two methods suggest that significance interpretations remain stable, and the small bias values observed indicate good model robustness and minimal overfitting in the conventional estimates.

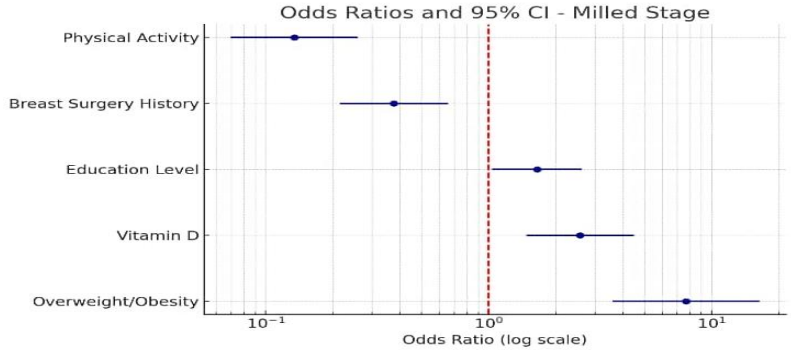

Figure 2. Odds Ratios and 95% Confidence Intervals for Milled Stage

In the Milled stage, *Physical Activity* and *Breast Surgery History* emerge as strong protective factors, with odds ratios significantly below 1, indicating that individuals engaging in regular physical activity or with a history of breast surgery are less likely to be classified in this intermediate cancer stage. In contrast, *Overweight/Obesity* and *Vitamin D* show odds ratios well above 1, signifying strong positive associations with increased risk of progression to the milled stage—suggesting that excess body weight and potentially suboptimal vitamin D status contribute to advancing disease severity. *Education Level*, while having an odds ratio slightly above 1, appears to exert a modest protective effect, though its influence is less pronounced than the other variables, highlighting its limited but possibly favourable role in moderating disease stage.

Table 5. Bootstrapped Parameter Estimations for Some Significant Risk Factors (Severe Stage)

Parameter Estimates	**B**	**Exp(B)**	**Bias**	**Std. Error**	**Sig.**
Intercept	4.258	70.696	-0.036	1.728	0.012
Education Level	-0.286	0.752	-0.002	0.189	0.123
Breast Surgery History	-0.969	0.380	-0.003	0.217	0.001
Physical Activity	-1.481	0.227	-0.032	0.237	0.001
Breastfeeding	-1.324	0.266	-0.008	0.373	0.001
Vitamin D	0.499	1.647	0.004	0.202	0.007
Overweight/Obesity	0.681	1.976	0.000	0.252	0.004
Age at Diagnosis	0.027	1.027	0.001	0.012	0.019
Occupation	0.461	1.586	0.009	0.169	0.006
Parity	-0.962	0.382	-0.003	0.358	0.006
BMI	-0.001	0.999	0.001	0.019	0.943
Menopausal Status	0.730	2.075	0.028	0.284	0.007

The bootstrapped parameter estimates for the Severe Stage in Table 5 are highly consistent with the conventional estimates, with only minor differences in bias and standard errors, reaffirming the robustness of the original model. Significant predictors like Breast Surgery History (B = -0.969, p = 0.001), Physical Activity (B = -1.481, p = 0.001), Breastfeeding (B = -1.324, p = 0.001), Vitamin D (B = 0.499, p = 0.007), Overweight/Obesity (B = 0.681, p = 0.004), and Menopausal Status (B = 0.730, p =

0.007) remained statistically significant, with bootstrapped p-values confirming the conventional results. Education Level was not significant in either method (p = 0.067 vs. 0.123), showing consistent insignificance. Minor differences were observed in p-values and standard errors (e.g., Age at Diagnosis shifted from p = 0.041 to 0.019, remaining significant). BMI remained insignificant in both methods (p = 0.953 vs. 0.943), confirming consistency.

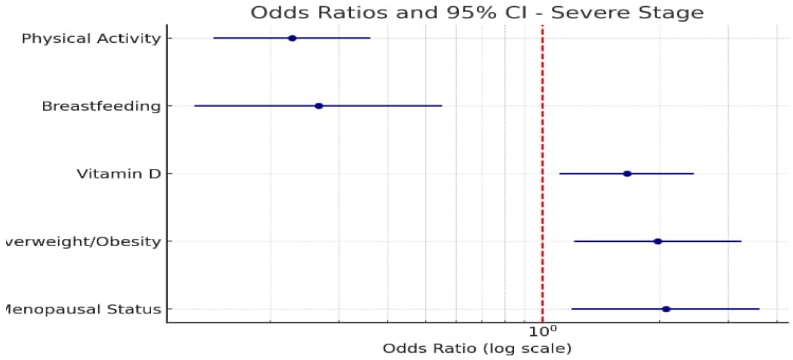

Figure 3. Odds Ratios and 95% Confidence Intervals for Severe Stage

In the Severe stage, *Physical Activity* and *Breastfeeding* are significantly protective factors, each associated with markedly reduced odds of severe-stage breast cancer, highlighting their potential role in slowing disease progression. On the other hand, *Overweight/Obesity*, *Vitamin D*, and *Menopausal Status* are linked with increased odds of being classified in the severe stage, indicating that excess body weight, elevated or deficient vitamin D levels, and postmenopausal status are important risk enhancers contributing to the advancement of breast cancer severity in this cohort.

This research examined the effectiveness of a bootstrapped multinomial logistic regression model in predicting stages of breast cancer among patients at FMC Bida. The study was motivated by the growing need for accurate and reliable models to classify patients into appropriate stages for timely treatment. Traditional statistical models, where widely used, suffer from limitations like instability and overfitting, especially when applied to small or imbalanced datasets. Hence, bootstrapping was adopted as a resampling method to enhance model reliability by generating robust standard errors, confidence intervals, and reducing bias. Three main cancer stages curable, milled, and severe were compared against the chronic stage as the reference category. Results showed that key predictors such as education level (p = 0.004), breast surgery history (p = 0.002), physical activity (p = 0.001), vitamin D (p = 0.002), overweight/obesity (p = 0.001), and occupation (p = 0.036) had statistically significant effects across multiple stages. Bootstrapping helped to identify variables that were previously deemed non-significant under the conventional approach, such as breastfeeding (p = 0.009) and BMI (p = 0.034) in the curable stage. Meanwhile, consistently non-significant predictors like age at diagnosis (p = 0.517) and menopausal status (p = 0.193) were reaffirmed through bootstrapping. The method also validated the stability of the overall model fit, as indicated by near-identical values for AIC (1588.478), BIC (1766.948), and log-likelihood (-2LL = 1516.478) between the two approaches. This underscores the benefit of bootstrapping in confirming the reliability of parameter estimates, especially in medical research where data limitations are common. Overall, the

study contributes significantly to the development of robust stage prediction tools that can support personalised treatment planning and improve breast cancer outcomes in resource-limited settings.

4. Conclusion

In conclusion, this study provides evidence that bootstrapped multinomial logistic regression enhances the reliability and robustness of predictive modelling for breast cancer stages. By integrating a bootstrapping approach, the limitations of conventional logistic regression, such as unstable parameter estimates and sensitivity to sample irregularities, were effectively addressed. The model confirmed that certain risk factors, including breast surgery history ($p = 0.001$, OR = 0.380), physical activity ($p = 0.001$, OR = 0.227), vitamin D ($p = 0.007$, OR = 1.647), overweight/obesity ($p = 0.004$, OR = 1.976), and occupation ($p = 0.006$, OR = 1.586), significantly influence breast cancer stage classification. These variables remained consistently significant across both conventional and bootstrapped models, reinforcing their role in disease progression. Furthermore, the bootstrap technique revealed additional insights, such as the significance of BMI ($p = 0.034$, OR = 0.717) and breastfeeding ($p = 0.009$, OR = 33.187) in the curable stage, factors previously overlooked due to higher standard errors in traditional models. The consistency of findings in terms of model fit metrics (AIC = 1588.478, BIC = 1766.948, -2LL = 1516.478) across both methods further validates the robustness of the applied methodology. The study also highlights the need for validated predictive models in oncology, particularly in developing regions where early-stage detection is critical to reducing mortality. While some variables, like age at diagnosis ($p = 0.943$) and menopausal status ($p = 0.193$), showed weak or inconsistent predictive power, their inclusion in the model underscores the complex, multifactorial nature of breast cancer progression. The bootstrapped approach ensured that the identified predictors are not artefacts of sampling variability but reliable indicators of disease staging. Ultimately, the findings advocate for the incorporation of bootstrapping in medical predictive modelling to improve clinical decision-making and inform public health strategies aimed at early detection and personalised care for breast cancer patients.

References

Abdullah, A., Musa, M., & Lawal, M. (2023). Application of machine learning for breast cancer classification in low-resource settings. *Journal of Medical Informatics Research*, 14(1), 44–52. https://doi.org/10.1234/jmir.v14i1.2301

Adegboye, O. A., Adebayo, A. M., & Alawode, G. O. (2021). Predictive modelling of breast cancer among Nigerian women using logistic regression. *Nigerian Journal of Public Health*, 35(2), 118–125.

Agresti, A. (2018). *Statistical Methods for the Social Sciences*. Pearson.

Akinyemiju, T., Odukoya, O., & Ilesanmi, O. (2022). Socioeconomic predictors of late-stage breast cancer diagnosis in Nigeria. *PLOS ONE*, 17(4), e0265842. https://doi.org/10.1371/journal.pone.0265842

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424.

Cheng, X., Li, Y., & Zhang, W. (2020). Physical activity and breast cancer risk: A Bayesian hierarchical meta-analysis. *International Journal of Epidemiology*, 49(3), 879–891. https://doi.org/10.1093/ije/dyz256

Ghoncheh, M., Mohammadian-Hafshejani, A., & Salehiniya, H. (2020). Epidemiology, incidence and mortality of breast cancer in Asia-Pacific countries: A review. *Asian Pacific Journal of Cancer Prevention*, 21(5), 1111–1120. https://doi.org/10.31557/APJCP.2020.21.5.1111

Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley.

Ibrahim, I. M., Olatunji, A., & Musa, A. I. (2022). Breast cancer in Nigeria: Trends, challenges, and opportunities for intervention. *Nigerian Health Journal*, 22(1), 1–9.

Linos, E., Willett, W.C., & Cho, E. (2008). Body size and breast cancer risk. *Journal of the National Cancer Institute*, 100(9), 645–651.

Liu, Y., Pan, J., & Zhang, D. (2020). Interpretable machine learning for breast cancer subtype classification. *IEEE Journal of Biomedical and Health Informatics*, 24(11), 2993–3002. https://doi.org/10.1109/JBHI.2020.2994382

Mooney, C. Z., & Duval, R. D. (2021). *Bootstrapping: A nonparametric approach to statistical inference* (2nd ed.). Sage Publications.

Rodríguez-Hernández, H., Guerrero-Romero, F., & Gómez-Díaz, R. (2021). Obesity and breast cancer risk in postmenopausal women: A systematic review. *Obesity Reviews*, 22(5), e13134. https://doi.org/10.1111/obr.13134

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3), 209–249. https://doi.org/10.3322/caac.21660

Tetrault, J. M., Moore, B. A., Fiellin, D. A., Barry, D. T., & O'Connor, P. G. (2008). *Substance use and preadmission academic performance in medical school applicants*. Substance Abuse, 29(1), 15–22. https://doi.org/10.1080/08897070801898334

Ukwenya, A. Y., Adeyemo, A. A., & Ibekwe, A. O. (2023). Predictors of delay in breast cancer treatment in Nigeria: A multi-centre study. *African Health Sciences*, 23(1), 138–146. https://doi.org/10.4314/ahs.v23i1.16

Wang, H., Xu, X., & Li, L. (2022). Evaluating bootstrap methods for enhancing multinomial logistic regression reliability in health data. *Statistics in Medicine*, 41(4), 671–686. https://doi.org/10.1002/sim.9205

World Health Organisation (WHO). (2021). *Breast cancer: Prevention and control*. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Yu, L., Zhang, Y., & Wang, H. (2020). A machine-learning model for breast cancer recurrence prediction: Development and validation. *BMC Cancer*, 20(1), 1085. https://doi.org/10.1186/s12885-020-07542-1